INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e de Gestão

Diagnóstico Inteligente de Patologias da Laringe

Nuno Filipe Ribeiro Alves

Dissertação apresentada à

Escola Superior de Tecnologia e Gestão

Instituto Politécnico de Bragança

para obtenção do grau de Mestre em

Tecnologia Biomédica

Orientador:

Professor Doutor João Paulo Teixeira

Bragança, outubro de 2016

AGRADECIMENTOS

Ao concluir esta tese resta-me agradecer a todas as pessoas que contribuíram para que esta fosse possível.

Ao orientador Professor Doutor João Paulo Teixeira pela paciência e conhecimentos transmitidos fundamentais no desenvolver deste trabalho e enriquecimento pessoal.

Á minha família em geral e a minha mãe em particular por todo o apoio e amor incondicional.

RESUMO

A deteção automática de patologias da laringe permite fazer um diagnóstico rápido, barato e de forma não invasiva. Ao longo deste trabalho foram estudados vários tipos de parâmetros, sistemas de inteligência artificial e técnicas de seleção de variáveis que possam permitir a deteção de patologias das cordas vocais.

Foram utilizados um primeiro conjunto de parâmetros constituídos por HNR e quatro medidas de jitter e shimmer. Foi avaliada a capacidade de predição deste conjunto de parâmetros quando usados com apenas uma vogal e um tom e quando usados com varias vogais e vários tons.

Foi estudado um segundo conjunto de parâmetros onde constam 12 coeficientes cepstrais, frequências e larguras de banda dos três primeiros formantes, frequência fundamental, energia, potencia, momentos espectrais de ordem zero, um, dois, três e curtose. Isto serviu para aferir a utilidade de outro tipo parâmetros na deteção de patologias da laringe.

Devido á grande quantidade de parâmetros e para melhor compreender a utilidade de alguns foram aplicadas técnicas de seleção de variáveis e redução da dimensão como a regressão linear passo a passo e análise das componentes principais (PCA).

Foram utilizados dois tipos de sistemas inteligentes que depois de treinados permitiam a classificação em patológico ou saudável, as redes neuronais artificiais (RNA) e máquinas de vetor de suporte (MVS).

Como grupos patológicos foram usadas a disfonia e paralisia das cordas vocais, separadas por género. Na classificação como patológico ou saudável, usando o primeiro conjunto de parâmetros (HNR, quatro medidas de jitter e shimmer para três vogais e três tons diferentes), foi possível obter precisões de: 100% usando tanto a disfonia feminino como masculino como grupo patológico; 78,9% usando a paralisia feminino como grupo patológico; 81,8% usando a paralisia masculino como grupo patológico.

Palavras-chave: Patologias da Laringe, Deteção Automática, Redes Neuronais Artificias, Máquinas de Vetor de Suporte, Seleção de Parâmetros, PCA.

ABSTRACT

Automatic detection of laryngeal pathologies allows a fast, low-cost and noninvasive diagnosis. Throughout this work we studied various types of parameters, artificial intelligence systems and variable selection techniques that can allow detection of pathologies of the vocal cords.

We used a first set of parameters consisting of HNR and four measures of jitter and shimmer. The prediction capacity of this set of parameters was evaluated when used with only one vowel and tone and when used with several vowels and tones.

We studied a second set of parameters which contains 12 cepstral coefficients, frequencies and bandwidths of the first three formants, fundamental frequency, energy, power, spectral moments of order zero, one, two, three and kurtosis. This served to assess the utility of other parameters in the detection of disorders of the larynx.

Due to the large size of input parameters and to better understand the usefulness of some, variable selection and dimension reduction techniques like linear stepwise regression and principal component analysis (PCA) were applied.

Two types of intelligent systems, like artificial neural networks (ANN) and support vector machine (MVS) that after training allows to classify in healthy or pathologic voices, were used.

Dysphonia and vocal cord paralysis, separated by gender, were used as pathological groups. In the classification as pathological or healthy, using the first set of parameters (HNR, four measures of jitter and shimmer for three vowels and three different tones), it was possible to obtain accuracies of: 100% using both female and male dysphonia as pathological group; 78.9% using the female paralysis as pathological group; 81.8% using male paralysis as pathological group.

Key words: Larynx Pathologies, Automatic Detection, Artificial Neural Networks, Support Vector Machine, Parameters Selection, PCA.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

ABREVIATURAS E SÍMBOLOS

Lista de abreviaturas

CF	Controlo Feminino
CM	Controlo Masculino
P40F	Disfonia Feminino
P40M	Disfonia Masculino
P136F	Paralisia das cordas vocais Feminino
P136M	Paralisia das cordas vocais Masculino
SVD	Saarbrucken Voice Database
RNA	Rede Neuronal Artificial
MVS	Máquina de Vetor de Suporte
FTCE	Função Transferência Camada Escondida
FTS	Função Transferência da Saída
FT	Função de Treino
MFCC	Mel Frequency Cepstral Coeficients
F0	Frequência Fundamental
F1	Frequência do primeiro formante
F2	Frequência do segundo formante
F3	Frequência do terceiro formante
Bw1	Largura de banda do primeiro formante
Bw2	Largura de banda do segundo formante
Bw3	Largura de banda do terceiro formante
M0	Momento espectral de ordem zero
M1	Momento espectral de ordem um
M2	Momento espectral de ordem dois
M3	Momento espectral de ordem três
K	Curtose
Arq.	Arquitetura
R-T	Valor de R do conjunto de teste
R-P2	Valor de R do conjunto de treino e validação
Prec.-T	Precisão do conjunto de teste
Prec.-P2	Precisão do conjunto de treino e validação

Nº neur.	Número de neurónios da camada escondida
Param.	Parâmetros
Mét.	Método
Espec.	Especificidade
Sens.	Sensibilidade
tansig	Função transferência tangente sigmóide
purelin	Função transferência linear
trainlm	Função treino Levenberg-Marquardt
trainscg	Função treino gradiente conjugado
logsig	Função transferência Log-sigmóide

CAPÍTULO I

1. INTRODUÇÃO

A comunicação oral é uma das mais importantes formas de expressão humana (Almeida, 2010).

Na produção de voz intervêm vários sistemas. O sistema respiratório, que é a fonte de energia, responsável pelo ar expelido pelos pulmões. O sistema fonatório, que é a fonte de vibrações, representado pelas pregas vocais. O sistema de ressonância que inclui a cavidade oral e nasal. O sistema articulatório do qual fazem parte a língua, lábios, mandibula, palato e dentes. E o sistema nervoso central e periférico, como o córtex, que permite a coordenação (Almeida, 2010). Contudo e de uma forma resumida podemos afirmar que a voz é gerada na laringe pela vibração das cordas vocais e depende de um fluxo de ar adequado dos pulmões. O som vozeado é depois filtrado, amplificado e modulado pelos órgãos do trato vocal para formar a fala (Brockmann-Bauser, 2011).

Qualquer perturbação na voz trará implicações profundas na vida social e profissional de uma pessoa (Almeida, 2010), (Lopes, 2008). Em pacientes com patologias progressivas é de suma importância ter acesso a um rápido diagnóstico a fim de promover um melhor tratamento e prognóstico (Pylypowich & Duff, 2016).

A rouquidão é uma das principais queixas ouvidas no dia-a-dia dos centros de saúde. A prevalência global de disfonia é de 30% nos adultos e 50% nos adultos mais velhos. A rouquidão é também conhecida como disfonia (Pylypowich & Duff, 2016). A disfonia é um termo médico que significa desordem (dis-) da voz (-fonia) (Teixeira & Fernandes, 2015). As patologias da voz são bastante comuns e afetam cerca de 5% da população (Mora et al, 2006). As pessoas que tenham ocupações em que a voz é o instrumento principal no cargo que desempenham, como professores, têm um risco acrescido de vir a desenvolver disfunções vocais (Pylypowich & Duff, 2016).

Embora existam vários exames que podem ser feitos para detetar patologias associadas á voz estes ou são de cariz invasivo (vídeo-laringoscopia) ou dependem da experiencia do médico que faz a avaliação (exame auditivo) (Brockmann-Bauser, 2011), (Teixeira & Fernandes, 2015). A taxa de acerto de um exame auditivo feito por um médico pode variar entre 60 e 70% (Uloza et al, 2010).

I-Introdução

A análise acústica da voz é uma técnica bastante utilizada na deteção e estudo de patologias da voz (Brockmann-Bauser, 2011). Correlaciona-se, em geral, com o uso de técnicas computacionais que visam medir propriedades do sinal acústico de uma voz gravada dizendo vogais de forma sustentada ou em discurso. Incluem medidas de frequência fundamental (F0), índices de frequência e perturbação da amplitude (jitter e shimmer), análise espectral, entre outros (Brockmann-Bauser, 2011).

Contudo, em grande parte dos casos a extração de parâmetros conduz a um grande número de variáveis. Esse elevado conjunto de variáveis aliada ao elevado número de exemplos que devem ser estudados para encontrar um padrão tornam a tarefa de classificação humanamente impossível. É aí que entram os classificadores inteligentes. Sistemas como as redes neuronais artificiais (RNA) e as máquinas de vetor de suporte (MVS) são dos mais utilizados. Existem também técnicas de redução da dimensão e seleção de variáveis que permitem otimizar os conjuntos de treino destes sistemas. Uma das técnicas mais conhecidas é a análise das componentes principais (PCA).

Os parâmetros acústicos aliados a classificadores inteligentes podem acabar com a subjetividade dos exames auditivos, melhorar a taxa de assertividade para além do benefício do carácter não invasivo.

Esta dissertação está organizada em sete capítulos. No primeiro capítulo é feita uma introdução ao tema, onde é referida a vantagem em utilizar a análise acústica vocal na deteção de patologias da laringe, seguida de uma revisão da literatura. O segundo capítulo versa sobre a base de dados e patologias utilizadas neste estudo, assim como uma breve descrição sobre estas. No capítulo três é feita uma introdução e descrição teórica de todos os parâmetros usados neste estudo. No capítulo quatro são descritas algumas ferramentas de inteligência artificial bem como métodos para selecionar parâmetros/variáveis. No capítulo cinco são referidos os algoritmos e ferramentas usadas na análise acústica vocal e extração de parâmetros, assim como a implementação de algumas ferramentas de inteligência artificial em ambiente Matlab. No capítulo seis são apresentados os resultados e discussão sobre algumas das experiencias feitas neste trabalho. No capítulo sete estão presentes as conclusões retiradas sobre todos os estudos feitos ao longo deste trabalho e algumas sugestões para trabalhos futuros.

1.1. ESTADO DA ARTE

Em Henríquez et al, (2009), é estudada a utilidade de seis medidas caóticas não lineares baseadas na teoria de dinâmica não-linear na discriminação entre dois níveis de qualidade de voz: saudável e patológica. As medidas estudadas são entropias de primeira e de segunda ordem Rényi, a entropia de correlação e a dimensão de correlação. Os valores do primeiro mínimo da função de informação mútua e entropia de Shannon também foram estudados. Duas bases de dados foram utilizadas para avaliar a utilidade das medidas: um banco de dados multi-qualidade composto por quatro níveis de qualidade de voz (voz saudável e três níveis de voz patológica); e um banco de dados comercial (MEEI) composto por dois níveis de qualidade de voz (vozes saudáveis e patológicos). Um classificador baseado em redes neuronais padrão foi implementado a fim de avaliar as medidas propostas. Foram obtidas taxas de sucesso global de 82,47% (base de dados multi-qualidade) e 99,69% (base de dados comercial).

Em Forero et al, (2015), são utilizados parâmetros do sinal glotal para classificação em três grupos diferentes: pacientes com nódulos nas cordas vocais, pacientes com paralisia unilateral das cordas vocais e pacientes com voz saudável. A fase de fecho (Ko), a fase de abertura (Ka), quociente de abertura (OQ), quociente de fecho (CIQ), quociente de amplitude (AQ), quociente de amplitude normalizada (NAQ), quociente de abertura calculado pelo modelo Liljencrants-Fant (OQa), quociente de quase abertura (QOQ), quociente de velocidade (SQ), diferenças entre harmónicos (DH12), fator que mede a riqueza em harmónicos (HRF), Jitter e Shimmer são os parâmetros extraídos com recurso ao *software* Aparat utilizados neste trabalho. A base de dados cedida por um terapeuta da fala contém 12 pacientes com nódulos, 8 com paralisia das cordas vocais e 11 saudáveis, com 8 gravações por paciente. Redes neuronais artificiais, máquinas de vetor de suporte e cadeias de Markov escondidas são os métodos de classificação empregados, permitindo assim uma taxa de acerto de 95,8, 82 e 96,2 % respetivamente.

Em Markaki & Stylianou (2011), são exploradas as informações fornecidas por uma representação referida como modelação espectral, para a deteção e discriminação dos distúrbios da voz. A representação inicial é primeiro transformada num domínio dimensional inferior usando a decomposição em valores singulares de ordem superior (HOSVD). A partir desta representação de menor dimensão é sugerido um processo de seleção de parâmetros baseado na informação mútua entre as classes de voz (ou seja, normofonicas / disfônicas).

Para avaliar a abordagem sugerida e representação, foram realizadas experiências, utilizando máquinas de vetor de suporte (MVSs) para a classificação. Para a deteção de patologias da voz, a abordagem sugerida alcançou uma precisão de classificação de 94,1%.

Em Eskidere & Gurhanh (2015), é usado um novo método para obter os coeficientes de mel-cepstral (MFCC). Este método consiste em usar um sistema de múltipla janela de Thomson em vez da tradicional janela única de Hamming. Este novo método provou ser melhor do que o anteriormente utilizado alcançando uma precisão média de 99,38 % em relação aos 95 % obtidos até então. A base de dados usada foi a Saarbruecken Voice Database (SVD) constituída por 650 vozes saudáveis e 650 vozes com as mais variadas patologias. O modelo de mistura gaussiana (GMM) foi o método de classificação escolhido.

Em Fezari et al, (2014), são usados como parâmetros os MFCC's em conjunto com Jitter e Shimmer, para a deteção de uma patologia chamada disfonia espasmódica. A base de dados usada é a Saarbrucken Voice Database (SVD) e o método de classificação é o Gaussian Mixture Model (GMM). Este trabalho é feito com base na frase "Bom dia, como está" dita em alemão. O objetivo em usar a frase, por um lado é obter mais dados para treino, onde o GMM precisa de uma quantidade importante de dados especialmente quando se usa um número elevado de mistura (Gaussiana), por outro, a diversidade de dados que pode aumentar a precisão de um sistema. No pré-processamento é feita a remoção de silêncios e selecionadas apenas as vogais /a/ e /u/. Um aspeto importante é o uso de várias *frames* por pessoa. Ficando a classificação depende de um valor de *threshold*. "Se mais de 70% das *frames* de um sinal forem atribuídos a determinada classe então assume-se que todo sinal pertence aquela classe." A melhor precisão obtida foi de 82,31%.

Em Malyska et al (2005), é usado um método para a modulação da amplitude (AM) banda de frequências. É projetado um sistema de reconhecimento da disfonia na voz usando um modelo biologico inspirado no colículo inferior para avaliar a performance deste método. O sistema é construído sobre o GMM como modelo de classificação e recorre á base de dados da Kay Elemetrics, a MEEI. A melhor performance foi alcançada com o uso do método proposto em conjunto com os MFFC's, com uma precisão de 95,6%.

Em Panek et al, (2015), um vetor composto de 28 parâmetros acústicos é avaliado usando análise das componentes principais (PCA), análise do *kernel* PCA (kPCA) e uma rede neuronal auto-associativa (NLPCA) na deteção de quatro tipos de patologia (disfonia hipertónica, disfonia funcional, laringite, paralisia das cordas vocais) usando as vogais /a/, /i/

e /u/, faladas em tom alto, baixo e normal. Os resultados indicam que os métodos kPCA e NLPCA podem ser considerados um passo para a deteção de patologias das cordas vocais. Os resultados mostram que esta abordagem proporciona resultados aceitáveis para este fim, com os melhores níveis de eficiência de cerca de 100%. De referir também que a classificação entre patológico e saudável foi feita de forma separada para cada doença e género, selecionando uma quantidade de pacientes de controlo igual ao número de pacientes de determinada patologia.

O trabalho desenvolvido em Al-Nasheri et al, (2016), concentra-se no desenvolvimento de um método robusto e preciso para extração de características do sinal para deteção e classificação de patologias da voz através da investigação de diferentes bandas de frequência usando as funções de correlação. Neste trabalho, foram extraídos o pico máximo e respetivo valor de atraso para cada janela do sinal, usando funções de correlação para detetar e classificar amostras patológicas. Essas características são investigadas em diferentes bandas de frequência para ver a contribuição de cada banda sobre os processos de deteção e classificação. Várias amostras de vozes normais e patológicas da vogal /a/ dita de forma sustentada foram extraídas a partir de três bases de dados diferentes: Arabic Voice Pathology Database (AVPD), Saarbruecken Voice Database (SVD) e Massachusetts Eye and Ear Infirmary (MEEI). Uma máquina de suporte de vetor foi utilizada como classificador. As melhores precisões alcançadas variaram de acordo com a banda, a função de correlação, e a base de dados. As bandas que mais contribuíram tanto na deteção como classificação foram entre 1000 e 8000 Hz. Na deteção, a precisão mais elevada foi alcançada usando correlação cruzada, 99,8%, 90,9% e 91,1% MEEI, SVD e AVPD, respetivamente. Contudo, na classificação, a precisão mais alta foi de 99,2%, 98,9% e 95,1% nos três bancos de dados, respetivamente.

Em Sellam & Jagadeesan, (2014), são explorados e comparados vários modelos de classificação para aferir a capacidade dos parâmetros acústicos em diferenciar vozes normais de vozes patológicas. É feita uma tentativa de analisar e discriminar voz patológica de voz normal em crianças, utilizando diferentes métodos de classificação. A classificação em voz patológica e voz normal é feita implementando uma Máquina de Vetor de Suporte (MVS) e uma Rede Neuronal com função de base radial (RBFNN). O sinal de voz é analisada para extrair os parâmetros acústicos, tais como a energia do sinal, *frequência fundamental*, frequências formantes, sinal residual quadrático médio, coeficientes de reflexão, jitter e shimmer. A base de dados continha gravações de 10 vozes saudáveis e 10 patológicas. Os

melhores resultados foram obtidos para a rede neuronal, 91%, tendo a máquina de vetor de suporte obtido 83%.

CAPÍTULO II

2. BASE DE DADOS

Nas experiências realizadas neste trabalho foi utilizada uma base de dados alemã, Saarbrucken Voice Database (SVD), disponibilizada online de forma gratuita pelo Instituto de Fonética da Universidade de Saarland. Esta base de dados contém sinais de voz de mais de 2000 sujeitos saudáveis e com patologia. Para cada sujeito é disponibilizada a gravação dos fonemas /a/, /i/ e /u/ nos tons baixo, normal e alto ditos de forma sustentada e ainda a gravação de uma frase em alemão: ‘‘Guten Morgen, wie geht es Ihnen?’’ (Bom dia, como estás?). O tamanho dos ficheiros situa-se entre 1 e 3 segundos. A frequência de amostragem dos sinais de voz é de 50 kHz (Teixeira & Gonçalves 2014).

2.1. SINAIS PATOLÓGICOS UTILIZADOS

Como em outros estudos (por ex. (Panek et al, 2015)) a classificação em saudável e patológico foi feita separando o género feminino do masculino. O número de pacientes saudáveis selecionados foi o mesmo que o grupo patológico em estudo. Mais detalhes sobre o número de pacientes e a distribuição de idades podem ser vistos na Tabela 1. As patologias utilizadas neste estudo foram a paralisia das cordas vocais e a disfonia por serem dois grupos patológicos com mais sinais disponíveis na base de dados.

Tabela 1-Distribuição de idades por género e patologia dos grupos selecionados da base de dados SVD.

	Pacientes		Margem (anos)		Média (anos)		Desvio padrão (anos)	
	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino
Paralisia	126	69	21-79	23-81	55,8	59,1	12,4	14,4
Disfonia	41	29	18-73	11-77	45,6	48,7	14,8	18,0
Controlo	126	69	18-84	18-69	31,0	34,8	15,9	15,8
Controlo	41	29	19-56	20-69	24,7	41,2	7,2	18,7

2.1.1. DISFONIA

Disfonia é um termo médico que significa desordem (dis-) da voz (-fonia) (Teixeira & Fernandes, 2015). A voz humana é originada pelo fluxo de ar que vem dos pulmões e passa pelas cordas vocais. Este som é diferente da fala, a qual é modulada pela faringe, língua e cavidade oral (Pylypowich & Duff, 2016). Embora existam muitas causas de disfonia (Teixeira & Fernandes, 2015), esta pode ser caracterizada como um distúrbio no mecanismo fonatório causando alterações na frequência fundamental (Frequência fundamental) (Pylypowich & Duff, 2016). Uma perturbação na voz não é uma doença por si só mas pode ser um sintoma de uma patologia subjacente (Pylypowich & Duff, 2016).

Basicamente a disfonia é um distúrbio na comunicação, caracterizado pela dificuldade na produção vocal, registando-se um impedimento na produção natural de voz. Pode ser causada por um disfunção, uso intensivo ou mau uso da voz, é mais frequente em indivíduos que usam a voz diariamente de forma abundante e incorrecta. As pessoas com esta patologia podem apresentar rouquidão, dor de garganta ou garganta seca como sintomas. Um cantor ou cantora pode notar que já não é mais capaz de cantar em tons mais altos. Pode haver outros sintomas associados, como um gotejamento contínuo na parte de trás da garganta (catarro nasal) e azia (Teixeira & Fernandes, 2015).

Existe uma relação entre saúde vocal, distúrbios de voz (disfonia) e condições de trabalho. A disfonia pode se manifestar através de uma série de mudanças: dificuldade em manter a voz; Fadiga vocal; Variações na frequência usual; rouquidão; Falta de volume e projeção; Perda de eficiência vocal e pouca resistência ao falar (Teixeira & Fernandes, 2015).

A disfonia é na verdade uma patologia que está afeta a vários distúrbios e sintomas, manifestando-se tanto como sintoma secundário como principal. A disfonia pode ser orgânica ou funcional. Disfonia orgânica é devido a uma alteração anatómica na prega vocal, como nódulos ou tumores benignos. Quando não existem alterações anatómicas conhecidas a disfonia é assumida como funcional. Entre estes casos pode considerar-se a disfonia funcional orgânica que é geralmente iniciada com uma disfonia funcional não tratada e progride para lesões secundárias da prega vocal (Teixeira & Fernandes, 2015).

2.1.2. PARALISIA DAS CORDAS VOCAIS

A paralisia das cordas vocais é um distúrbio da voz que ocorre quando uma (unilateral) ou ambas (bilateral) pregas vocais não abrem ou fecham de forma apropriada. A paralisia unilateral é distúrbio comum, enquanto a bilateral é mais rara e pode implicar risco de vida.

As cordas vocais são duas bandas elásticas presentes na laringe logo acima da traqueia. Aquando da respiração estas permanecem afastadas e na deglutição elas ficam fechadas. Contudo, na produção de voz o ar que vem dos pulmões faz com que estas vibrem oscilando entre a posição aberta e fechada.

Em casos de paralisia, as cordais vocais podem permanecer abertas deixando as vias respiratórias e pulmões desprotegidos. Este tipo de patologia tanto pode ocorrer após trauma na cabeça, pescoço ou peito como em pessoas com problemas neurológicos como esclerose múltipla, doença de Parkinson ou que tenham sofrido um AVC (acidente vascular cerebral).

Os sintomas podem manifestar-se sobre a forma de rouquidão, soprosidade, dificuldades em respirar, respiração ruidosa e problemas de deglutição. Podem ainda ocorrer alterações na qualidade de voz como a perda de volume ou frequência fundamental (U.S. Department of Health & Human Services, 2011).

CAPÍTULO III

3. PARÂMETROS EXTRAÍDOS DO SINAL ACÚSTICO

3.1. INTRODUÇÃO

Nesta secção são descritos todos os parâmetros utilizados no estudo realizado. Do primeiro conjunto de parâmetros fazem parte o Jitter, Shimmer e HNR extraídos pelo algoritmo desenvolvido por (Teixeira & Gonçalves, 2016). Na figura 1 podemos ver uma ilustração do conceito de Jitter e Shimmer.

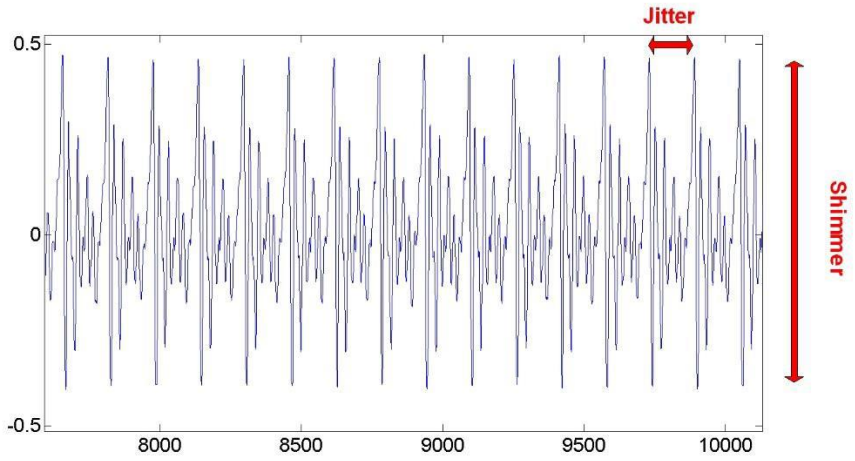

Figura 1-Representação do Jitter e Shimmer num sinal de voz (Teixeira & Gonçalves, 2014).

Do segundo conjunto de parâmetros constam coeficientes mel cepstrais (MFCC), frequências e larguras de banda dos três primeiros formantes, frequência fundamental, energia, potência, momentos espectrais de ordem zero, um, dois, três e curtose.

3.1.1. JITTER

O Jitter é definido como uma medida de variação do período glotal entre ciclos de vibração das pregas vocais (Teixeira & Fernandes, 2015). Os sujeitos que não consiguem controlar a vibração das cordas vocais têm tendência a ter valores de Jitter mais elevados. O jitter pode ser medido de quatro formas diferentes. Como absoluto, relativo, perturbação média relativa (relative average perturbation-rap) e o quociente de perturbação do período num intervalo de cinco pontos (five-points period perturbation quotient-ppq5).

Jitter absoluto é a variação da frequência fundamental entre ciclos, ou seja, a diferença absoluta média entre períodos consecutivos, expresso pela eq.1

$$jitta = \frac{1}{N-1}\sum_{i=1}^{N-1} |T_i - T_{i-1}| \tag{1}$$

Jitter relativo ou local é a diferença absoluta média entre períodos consecutivos, dividida pelo período médio e é expresso em percentagem (eq.2).

$$jitter(relative) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{2}$$

Jitter (rap) ou perturbação média relativa (relative average perturbation) é a diferença absoluta média entre um período e a média desse e os seus dois vizinhos, dividida pelo período médio. É expresso em percentagem e apresenta-se pela eq.3.

$$rap = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} |T_i - \frac{1}{3}\sum_{n=i-1}^{i+1} T_n|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{3}$$

Jitter (ppq5) ou quociente de perturbação do período num intervalo de cinco pontos (five-points period perturbation quotient-ppq5) é a diferença absoluta média entre um período e a

média desse e os seus quatro vizinhos dividida pelo período médio. É também expresso em percentagem (eq.4).

$$ppq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}|T_i - \frac{1}{5}\sum_{n=i-2}^{i+2} T_n\ |}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \quad (4)$$

Onde T_i é o tamanho do período glotal e N é número de períodos glotais.

3.1.2. SHIMMER

O Shimmer relaciona-se com a variação da amplitude a cada ciclo. Uma redução na resistência glotal e lesões podem causar variações da amplitude glotal correlacionadas com a soprosidade e emissão de ruído, dando lugar o valor de shimmer mais elevado. O Shimmer pode ser medido de quatro formas diferentes. Como absoluto em dB, relativo, quociente de perturbação da amplitude em três pontos (three point amplitude perturbation quotient-apq3) e quociente de perturbação da amplitude em cinco pontos (five point amplitude perturbation quotient-apq5) (Teixeira & Fernandes, 2015).

Shimmer absoluto (dB) expresso como a variação da amplitude pico a pico em decibel, ou seja, é o logaritmo de base 10 da média absoluta da razão da amplitude entre períodos consecutivos multiplicada por 20. É expresso em decibel (eq.5).

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20 * log\left(\frac{A_{i+1}}{A_i}\right)\right| \quad (5)$$

Shimmer relativo é definido como a diferença absoluta média entre amplitudes de períodos consecutivos, dividida pela amplitude média, expresso em percentagem (eq.6).

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \quad (6)$$

Shimmer (apq3) ou o quociente de perturbação da amplitude em três pontos (three point amplitude perturbation quotient-apq3) é a diferença absoluta média entre a amplitude de um período e a média das amplitudes dos seus vizinhos, dividida pela amplitude média. É expresso em percentagem (eq.7).

$$apq3 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1} A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \quad (7)$$

Shimmer (apq5) ou o quociente de perturbação da amplitude em cinco pontos (five point amplitude perturbation quotient-apq5) é a diferença absoluta média entre a amplitude de um período e a média das amplitudes dos seus quatro vizinhos, dividida pela amplitude média. É também expresso em percentagem (eq.8).

$$apq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|A_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2} A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \quad (8)$$

Onde A_i é a amplitude pico a pico e N é o número de períodos.

3.1.3. HNR

A relação entre as componentes harmónicas e de ruido (**Harmonic to Noise Ratio –HNR**) fornece uma indicação da periodicidade global do sinal de voz pela quantificação da relação entre a componente periódica (parte harmónica) e aperiódica (ruido). Este parâmetro é medido como uma característica geral do sinal, e não como uma função da frequência. O valor global de HNR de um sinal varia porque diferentes configurações do trato vocal implicam diferentes amplitudes para os harmónicos. O valor de HNR pode ser determinado pela eq.9 (Teixeira & Fernandes, 2015).

$$HNR = 10 * log_{10} \frac{AC_V(T)}{AC_V(0) - AC_V(T)} \quad (9)$$

Onde $AC_V(T)$ representa a potência da componente harmónica do sinal e $AC_V(0)$ corresponde á potência total do sinal. A diferença das duas é assumida como sendo a componente de ruído.

3.1.4. MFCC

Os Coeficientes Cepstrais na Frequencia Mel, do Inglês Mel Frequency Cepstral Coefficients (MFCC), são parâmetros de curto termo baseados no espectro (Logan, 2000). Os MFCC's são baseados no ouvido humano para o qual a perceção das frequências não segue uma escala linear (Logan, 2000), (Tiwari, 2010). Foi então criada uma escala Mel, segundo a qual os parâmetros de MFCC se regem, e que utiliza um filtro linear para frequências abaixo dos 1000 Hz e logarítmico acima de 1 kHz (Logan, 2000), (Muda et al, 2010). Para o cálculo destes parâmetros é necessário seguir uma série de passos como podemos ver na Figura 2.

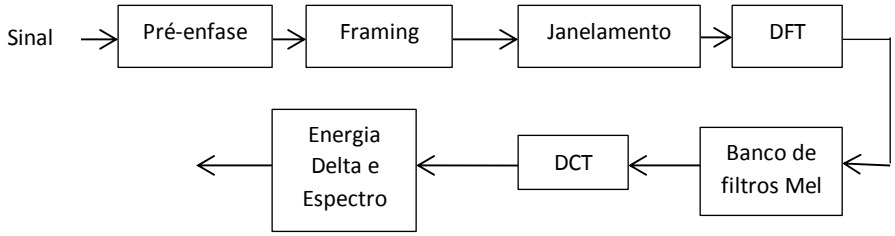

Figura 2-Diagrama dos passos a seguir para extração dos parâmetros MFCC.

Em primeiro lugar o sinal, aqui representado por *x[n]*, é filtrado por forma a realçar as frequências mais altas. Este processo irá aumentar a energia do sinal nas altas frequências (Muda et al, 2010). Este passo é designado de pré-enfase e está representado na equação 10.

$$y[n] = x[n] - a\, x[n-1] \quad (10)$$

Onde *y[n]* é o sinal depois de filtrado e o valor de a, usando 0,95, significa que presume-se que 95% de qualquer amostra é originada a partir da amostra anterior (Muda et al, 2010).

De seguida divide-se o sinal em *N* janelas, designadas de *frames*, com tamanhos a variar entre os 20 e os 40 ms. Aplica-se uma janela de Hamming de acordo com a equação 11. Onde *w[n]* representa a janela de Hamming.

$$w[n] = 0{,}54 - 0{,}46\cos\left(\frac{2\Pi n}{N-1}\right), 0 \leq n \geq N-1 \quad (11)$$

Em seguida faz-se a conversão do domínio dos tempos para o domínio das frequências através da Transformada de Fourier (FFT). Para cada *frame* calcula-se o periodograma da potência espectral. Como o espectro tem uma gama de valores muito alargada e o sinal de voz não segue uma escala linear é então aplicado o banco de filtros de acordo com a escala Mel. São usados filtros triangulares que servem para calcular uma soma ponderada das componentes espectrais de modo a que a saída se aproxime á escala Mel. A magnitude de cada filtro é igual a 1 no centro e decresce de forma linear até 0 nas pontas. A título de exemplo, se pretendermos um banco com 10 filtros significa que vamos ter 12 pontos igualmente espaçados sendo o mínimo o valor da frequência mínima do espectro e o máximo o valor de frequência máxima do espectro.

Depois a equação 12 é usada para converter a frequência em Hz para Mel (Molau et 2001).

$$F(Mel) = 2595 * log10\left[\frac{1+f}{700}\right] \quad (12)$$

A transformada discreta do cosseno, do Inglês discrete cosine transform (DCT), permite converter o espectro na base log Mel para o domínio dos tempos. O resultado da conversão é chamado de Coeficiente de Cepstro na Frequência Mel. O conjunto dos coeficientes é designado de vetores acústicos.

O último passo envolve o cálculo da energia e um fator designado de delta e pretende representar a dinâmica do sinal de *frame* para *frame*. Assim são adicionados aos 12 coeficientes de cepstro a energia, perfazendo 13 coeficientes delta ou de velocidade. Assim como 39 coeficientes duplo delta ou de aceleração. A Energia numa *frame* do sinal *x* de uma amostra no tempo *n1* para o tempo *n2* é expressa pela equação 13 (Muda et al, 2010).

$$Energia = \sum_{n=n1}^{n2} x^2(n) \quad (13)$$

Os coeficientes delta e duplo delta são também conhecidos como coeficientes diferenciais e de aceleração. Os coeficientes MFCC representam apenas a potência espectral de um único *frame* mas o sinal também contém informação na sua dinâmica, ou seja, quais são as trajetórias dos coeficientes MFCC ao longo do tempo. Como é sabido, calculando as trajetórias e juntando-as aos coeficientes de MFCC pode aumentar a performance de um sistema de análise acústica. Como tal teríamos 13 coeficientes MFCC mais 13 delta e 13 duplo delta perfazendo 39 coeficientes como estão referidos em cima. Cada um dos 13 coeficientes delta representa a variação de *frame* para *frame*. Para calcular os coeficientes delta usa-se a equação 14 (Muda et al, 2010).

$$d(t) = \frac{c(t+1) - c(t-1)}{2} \tag{14}$$

Onde d(t) representa o coeficiente delta da *frame* t calculado em termos de coeficientes estáticos c(t+1) e c(t-1). Os coeficientes duplo delta são calculados da mesma forma só que a partir dos delta e não dos coeficientes estáticos.

3.1.5. FORMANTES

Á medida que o fluxo de ar passa pelas cavidades acima da laringe, (faringe e boca), vão ser criadas ressonâncias a determinadas frequências. Estas frequências de ressonância vão determinar ou formar o espectro da onda sonora e são chamadas de formantes (Catford, 2001), (Schwarz, 1998). A boca e a faringe, responsáveis pelos ressoadores, mudam a sua configuração para cada vogal acentuando determinadas frequências características de determinada vogal. As frequências dos formantes são numeradas por ordem crescente de frequência, Formante 1, Formante 2, etc., sendo normalmente abreviadas para F1, F2, etc (Catford, 2001).

O envelope espectral contém informação sobre as frequências e larguras de banda dos formantes (Cordeiro et al, 2013) e deriva do espectro calculado pela Transformada de Fourier (Schwarz, 1998). Para achar o envelope espectral pode ser usado um método chamado de LPC (Linear Predictive Coding). A ideia do LPC é a de predizer cada amostra do sinal *s(n)* no domínio dos tempos por uma combinação linear dos *p* valores precedentes *s(n-p-1)* através de *s(n-1)*, *p* é chamado de ordem do LPC (Schwarz, 1998). Quanto maior for a ordem do LPC mais precisa vai ser a interpolação do espectro.

O valor aproximado de $\hat{s}(n)$ é calculado a partir dos p valores precedentes e coeficientes de predição a_i da seguinte forma:

$$\hat{s}(n) = \sum_{i=1}^{p} a_i \, s(n-i) \tag{15}$$

Para cada frame (janela) os coeficientes a_i vão ser calculados por forma a que o erro *e(n)*= $\hat{s}(n)$-*s(n)* seja mínimo.

Existe um filtro de análise dado pela função de transferência (eq.16), que tenta suprimir as frequências mais altas por forma a tornar o espectro mais achatado.

$$A(z) = 1 - \sum_{i=1}^{p} a_i z^{-1} \tag{16}$$

Por outro lado um filtro inverso, chamado de filtro de síntese dado pela eq.17, amplifica as frequências que foram atenuadas pelo filtro de análise.

$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-1}} \tag{17}$$

Como podemos observar o filtro de síntese, 1/A(z), é um filtro só com polos, uma vez que a sua função de transferência é definida por uma função racional sem zeros no numerador mas com p zeros no denominador A(z). Devido a estes zeros aparecerem em pares conjugados complexos, o valor absoluto da função de transferência (magnitude) do filtro resultante apresenta p/2 pólos ou picos. Á medida que a ordem do LPC decresce (isto é, menos pólos estão disponíveis) a aproximação do envelope ao espectro torna-se mais grosseira. Para o cálculo dos coeficientes de predição existem dois métodos: covariância e auto correlação (Schwarz, 1998).

3.1.6. MOMENTOS ESPECTRAIS

O método dos momentos providência uma técnica robusta para decompor uma forma arbitrária num conjunto de parâmetros característicos. Em geral, os momentos descrevem uma distância a determinado ponto ou eixo por quantidades numéricas. Os momentos são frequentemente utilizados em estatística para caracterizar distribuições e em mecânica para caracterizar a distribuição de massa dos corpos (Fujinaga, 1996).

Os momentos espectrais obtêm informação diretamente do espectro calculado pela transformada de Fourier (FFT). Uma das vantagens é a insensibilidade a mudanças de fase no sinal. Podem também ser vistos como um tipo de análise estatística feita ao espectro de potência (Vogel et al, 2001).

O cálculo dos momentos espectrais está por vezes limitado a determinada gama de frequências conhecendo á priori a física do espectro gerado. Isto corresponde a um processo de filtragem onde as frequências de interesse são extraídas (Vogel et al, 2001).

Passando do domínio dos tempos para o domínio das frequências a forma do espectro do sinal pode ser definida pelo momento espectral de ordem zero pela eq.18.

$$M_0(t) = \sum_{i=0}^{\infty} G(t, f_i) \tag{18}$$

Onde *G(t,f)* representa o espectro considerando a frequência central da i-ésima banda salientada na análise das frequências (Panek et al, 2015). O momento espectral de ordem zero é basicamente a média dos valores no intervalo definido e é proporcional á energia média nesse intervalo (Vogel et al, 2001).

O momento espectral de primeira ordem é o centro de gravidade do espectro (média ponderada da frequência) (Panek et al, 2015). Quando se trabalha com partes do espectro é necessário fazer a normalização com M_0 antes de extrair os momentos superiores (Vogel et al, 2001), daí a divisão por M_0 como podemos observar na eq.19.

$$M_1(t) = \frac{\sum_{i=0}^{\infty} G(t, f_i) f_i}{M_0(t)} \tag{19}$$

O momento espectral de segunda ordem pode ser interpretado como a variância da potência espectral (Vogel et al, 2001), ou o quadrado da largura espectral eq.20 (Panek et al, 2015).

$$M_2(t) = \frac{\sum_{i=0}^{\infty} G(t, f_i)[f_i - M_1(t)]^2}{M_0(t)} \tag{20}$$

O momento espectral de terceira ordem é descrito como a assimetria do espectro e é dado pela eq.21

$$M_3(t) = \frac{\sum_{i=0}^{\infty} G(t, f_i)[f_i - M_1(t)]^3}{M_0(t)} \tag{21}$$

O momento espectral de quarta ordem é também conhecido como curtose e mede o achatamento do espectro (eq.22) (Panek et al, 2015).

$$Curtose = \frac{M_4(t)}{M_2(t)^2} \tag{22}$$

Os momentos espectrais podem também ser calculados a partir do gráfico da Densidade da Potência Espectral, ou em Inglês Power Spectral Density (PSD) pela eq.23 (Sweitzer et al, 2004).

$$m_n = \int_0^{\infty} f^n G(f) df \tag{23}$$

Sendo que G(f) representa a potência em função da frequência e n é ordem do momento.

Ou na forma discreta pela eq.24.

$$m_n = \sum_{0}^{\infty} f^n G(f) \tag{24}$$

3.1.7. ENERGIA

A energia não é mais do que o somatório do sinal *x* no momento *n* ao quadrado, entre os intervalos de tempo n_1 e n_2, eq.25 (Panek et al, 2015).

$$E_x = \sum_{n=n1}^{n2} x^2(n) \tag{25}$$

3.1.8. FREQUÊNCIA FUNDAMENTAL (F0)

A, frequência fundamental ou F0 é o parâmetro físico resultante da vibração das pregas vocais por unidade de tempo (Lopes, 2008), (Sellam & Jagadeesan, 2014). A frequência fundamental permite avaliar a eficiência do sistema fonatório, a biomecânica laríngea bem como a sua interacção com a aerodinâmica (Lopes, 2008).

Este parâmetro depende de vários fatores: idade, sexo e comportamento vocal (Lopes, 2008).

O cálculo deste parâmetro pode ser feito no domínio dos tempos com recurso ao método da auto correlação. É feito diretamente no sinal (Sellam & Jagadeesan, 2014) e pressupõe o uso de técnicas de análise de curto termo (Tan & Karnjanadecha, 2003).

É comum fazer-se uma estimativa do valor de *frequência fundamental* achando o máximo da função de auto correlação (Tan & Karnjanadecha, 2003). Dado um sinal discreto *x(n)* a função de auto correlação é definida pela eq.26 (Sellam & Jagadeesan, 2014), (Tan & Karnjanadecha, 2003).

$$R_x(m) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n)x(n+m) \tag{26}$$

A função de auto correlação é basicamente uma transformação (não reversível) do sinal sendo útil para representar a estrutura da onda. Assim, para o cálculo do *frequência fundamental* se assumirmos que *x(n)* é periódico com período *P*, isto é *x(n)=x(n+P)* para todo o n, então a função de auto correlação é também periódica com o mesmo período, eq.27 (Sellam & Jagadeesan, 2014), (Tan & Karnjanadecha, 2003).

$$R_x(m) = R_x(m + P) \tag{27}$$

3.1.9. POTÊNCIA ROOT MEAN SQUARE (RMS)

A raiz quadrática média, do Inglês Root Mean Square (RMS), é um parâmetro que mede a ou potência do sinal e a sua forma discreta é dada pela eq.28 Adaptada de (Poomjan et al, 2014). Onde N representa o tamanho do sinal e *x(n)* o sinal.

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2(n)} \tag{28}$$

No algoritmo desenvolvido este parâmetro foi calculado a partir do sinal no domínio das frequências usando a FFT. Foi usada uma função do Matlab chamada *norm* que faz uma estimativa da energia do sinal discreto podendo ser representada pela eq.29 Onde *x* é o sinal e *N* o comprimento do sinal.

$$RMS = \frac{norm(x)}{\sqrt{N}} \tag{29}$$

CAPÍTULO IV

4. FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL

A análise de um grande conjunto de dados com múltiplas variáveis ou leva a que sejam cometidos erros ou é por vezes uma tarefa incomportável para um ser humano. Os sistemas de inteligência artificial são uma mais-valia e podem ser usados em tarefas de classificação (Kotsiantis, 2007). Um problema de classificação surge quando é necessário atribuir um objeto a uma classe ou grupo baseado num determinado número de parâmetros relacionados com o objeto (Zhang, 2000). Após treino, um sistema de inteligência artificial deve ter a capacidade para generalizar, ou seja, perante uma situação nunca antes vista tomar uma decisão com base em similaridades de parâmetros vistos anteriormente (Lanc, 1992).

4.1. REDES NEURONAIS ARTIFICIAIS (RNA)

O desenvolvimento de ferramentas computacionais capazes de realizar tarefas cognitivas que só são realizadas pelo cérebro humano é o foco principal da disciplina de Inteligência Artificial. As redes neuronais visam mimetizar o funcionamento do cérebro humano (Cruz, 2007). As redes neuronais artificiais (RNA) são sistemas simplificados do sistema nervoso central que podem ser implementadas por *software* ou *hardware* e são capazes de realizar tarefas (classificação ou regressão) após um período de treino (Cruz, 2007).

Uma boa definição para uma rede neuronal artificial talvez seja a de Robert Hecht-Nielsen, que as descreve como "estruturas de processamento de informação distribuídas em paralelo" (Lanc, 1992). As RNA são normalmente representadas por um diagrama composto por nós (neurónios) e ligações entre esse nós (sinapses) (Cruz, 2007), (Moraes et al, 2013), (Salhi et al, 2010). Os nós estão dispostos por camadas e a estrutura mais comum consiste em três camadas: camada de entrada (input layer), camada escondida (hidden layer) e a camada de saída (output layer). Uma rede com várias camadas é uma rede MLP (Multi-layer Perceptron) (Bishop, 1995). É também classificada como uma rede *feedforward* devido aos neurónios estarem conectados em apenas uma direção. Cada conexão tem um peso associado cujo valor é calculado pela minimização de uma função de erro global num processo de treino de gradiente descendente. Um neurónio é um modelo matemático simples que produz um valor de saída em dois passos. Primeiro, o neurónio calcula uma soma ponderada das entradas e

depois aplica uma função de ativação á soma por forma a criar um valor de saída. Habitualmente a função de ativação é não linear (Moraes et al, 2013) figura 3.

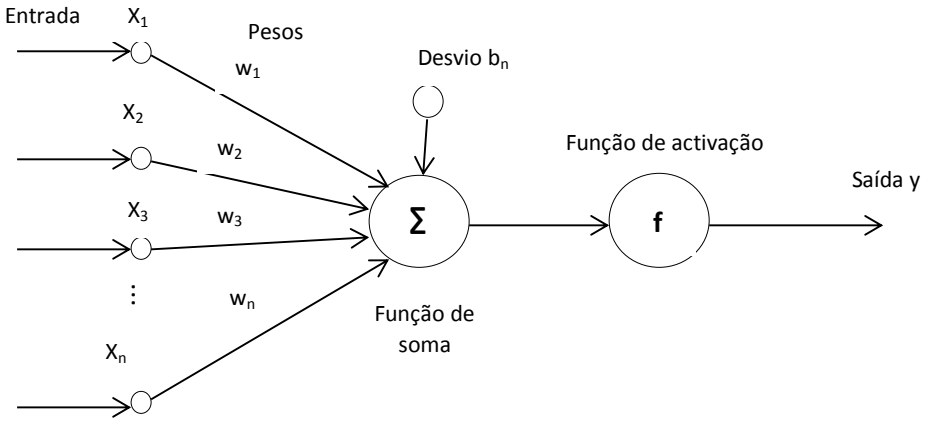

Figura 3-Diagrama representativo de uma rede neuronal artificial.

A rede aprende através de exemplos dados durante o treino e deve ser capaz de generalizar depois de treinada. A generalização é a capacidade de estimar determinadas características de um fenómeno nunca antes visto baseando-se em similaridades com parâmetros vistos anteriormente (Lanc, 1992).

4.1.1. TREINO DA RNA

Treinar uma RNA envolve a apresentação de um conjunto de padrões (parâmetros de entrada), calcular a saída (output) e compará-la com o valor desejado (target). Quando são apresentados os vetores de entrada á rede esta calcula um valor de saída que depois é comparado com o valor desejado. Os pesos são adaptados de forma sistemática otimizando a performance da rede. O processo de apresentação de exemplos e ajuste de pesos é repetido até a rede alcançar a performance desejada. Uma medida de performance da rede usual é a raíz do erro quadrático médio(Root Mean Square Error-RMSE) entre os valores desejados e os valores de saída. É desejável diminuir o RMSE, ou custo, ao mínimo possível. Contudo, usando métodos de gradiente descendente e o algoritmo *backpropagation* a RNA tende a ficar estagnada num mínimo local. Se a saída de um nó é relativamente mais significante como entrada do nó da camada seguinte, é atribuído um peso maior á saída desse nó. O esquema mais comum de treino é o de propagação do erro para trás, para os nós anteriores (algoritmo *backpropagation*) (Lanc, 1992). Contudo existem várias limitações no uso de uma RNA, tais como, ter de realizar varias experiencias para determinar a melhor estrutura e parâmetros da rede (Zekic-Susac, 2013), necessidade de ter um elevado numero de exemplos de treino, etc.

4.2. MÁQUINAS DE VETOR DE SUPORTE (MVS)

Uma máquina de vetor de suporte é um tipo de ferramenta inteligente baseada na minimização do risco estrutural. Podem ser usadas na resolução de problemas de classificação e de regressão. A ideia principal da MVS é a de construir hiperplanos como superfície de separação ótima entre exemplos positivos e negativos num contexto de classificação binária (Almeida, 2010; Sellam & Jagadeesan, 2014).

O conceito de "vetores de suporte" advém do "suporte" do algoritmo em alguns dados para estabelecer distâncias entre as classes. A figura 4 ilustra o conceito onde está representado um conjunto de dados do tipo (x_i; y_i) com x_i ϵR^n e y_i ϵ{-1;+1} (Cruz, 2007). Assim y tem o valor de -1 ou +1 de acordo com a classe a que x pertence (Almeida, 2010).

Na figura 4 podemos ver um classificador linear. A linha na diagonal representa a fronteira ou o hiperplano de separação das duas classes. A verde temos os exemplos positivos e a vermelho os negativos. Os vetores de suporte ou margens estão assinalados por uma circunferência.

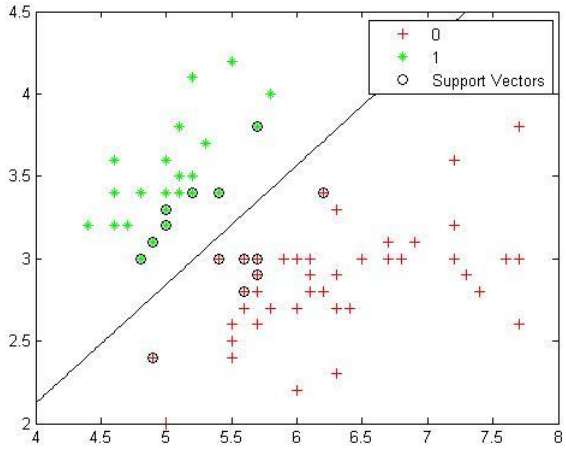

Figura 4-Classificador linear.

Uma linha de separação é do tipo:

$$w.x_i + b = 0 \tag{30}$$

Onde w é a inclinação da linha, x é o vetor de entrada e b a abcissa de início da linha (Almeida, 2010), (Cruz, 2007).

Existem várias linhas que podem servir de linhas de separação desde que respeitem as seguintes condições:

$$(w.x_i) + b \geq +1 \; se \; y_i = +1 \tag{31}$$

$$(w.x_i) + b \geq -1 \; se \; y_i = -1 \tag{32}$$

Contudo, importa maximizar a distância de separação aos dados bem como satisfazer ao mesmo tempo as condições supracitadas. Assim, a maior distância é determinada pela minimização dos vetores normais á linha, eq.33.

$$d(w,b) = min_{x_i|y_i} \frac{w.x_i + b}{|w|} - max_{x_i|y_i} \frac{w.x_i + b}{|w|} \tag{33}$$

Obtendo-se:

$$d(w,b) = min_{x_i|y_i} \frac{1}{|w|} - max_{x_i|y_i} -\frac{1}{|w|} = \frac{1}{|w|} - \frac{-1}{|w|} = \frac{2}{|w|} \tag{34}$$

Cuja derivada é:

$$d'(w) = \frac{1}{2}.|w| \tag{35}$$

A linha de separação que respeita as varias condições pode ser obtida pela seguinte função Lagrangeana, em que α_i são os multiplicadores Lagrangeanos:

$$L(w,b,\alpha) = \frac{1}{2}.|w|^2 - \sum_{i=l}^{n} \alpha_i . \{[(x_i.w) + b].y_i - 1\} \tag{36}$$

Esta função tem que ser minimizada em ordem a w e b, e maximizada em ordem a $\alpha_i \geq 0$, tendo um ponto óptimo a que corresponderá as soluções w_0, b_0 e α_i^0, resultando numa linha de separação com as seguintes propriedades:

$$\sum_{i=l}^{n} \alpha_i^0 . y_i = 0, \alpha_i^0 \geq 0, i = 1, \dots, n \tag{37}$$

$$w_0 = \sum_{i=l}^{n} \alpha_i^0 . y_i . x_i = 0, \alpha_i^0 \geq 0, i = 1, \dots, n \tag{38}$$

Segundo o teorema de Kunh-Tucker e substituindo na função Lagrangeana a solução é da forma de um vetor expresso sob a forma $\alpha^0 = \{\alpha_1^0, \ldots., \alpha_i^0\}$ podendo escrever-se a função de decisão:

$$f(x) = sign\left(\sum \alpha_i^0 . y_i . (x_i . x) - b_0\right) \tag{39}$$

Sendo x_i os vectores de suporte e $b_0 = \frac{1}{2}\left[\left(w_0 . x^*(1)\right) + \left(w_0 x^*(-1)\right)\right]$, com $x^*(1)$ um qualquer vector de suporte pertencente á primeira classe e $x^*(-1)$ um qualquer vector de suporte pertencente á segunda classe.

Uma vez que os problemas nem sempre são lineares é necessário fazer uma transformação dos dados para que estes possam ser separados linearmente. Para essa separação as MVS recorrem a métodos de *Kernel* que fazem uma transformação não linear aos dados para um espaço multi-dimensional onde ficará uma imagem dos dados que permita uma separação linear (Cruz, 2007).

Entre os métodos de *kernel* mais utilizados encontram-se o linear, polinomial, *radial basis function* (RBF) e *multi layer perceptron* (MLP) (Cruz, 2007). No treino de uma MVS são ajustados os parâmetros α_i e b para que a distância do hiperplano aos dados seja máxima. A MVS tem ainda outro conjunto de parâmetros designados de híper-parâmetros dos quais a função *kernel* está dependente como a constante C das linhas de fronteira que ladeiam o hiperplano, a largura do *kernel* Gaussiano e o grau do *kernel* polinomial, entre outros (Ben-Hur & Weston, 2010).

A escolha do *kernel* pode determinar-se importante para o sucesso da MVS (Cruz, 2007). No caso dos híper-parâmetros acima referidos podem ser enunciadas algumas características relativas aos valores de cada *kernel*. No caso de um *kernel* linear um valor de C mais baixo permite ignorar pontos mais próximos da fronteira (hiperplano que separa as duas classes) aumentando a margem (as margens são definidas pelos dados (Cruz, 2007)). Para um valor de C mais alto as margens tornam-se mais próximas da fronteira (Ben-Hur & Weston, 2010), fig.5.

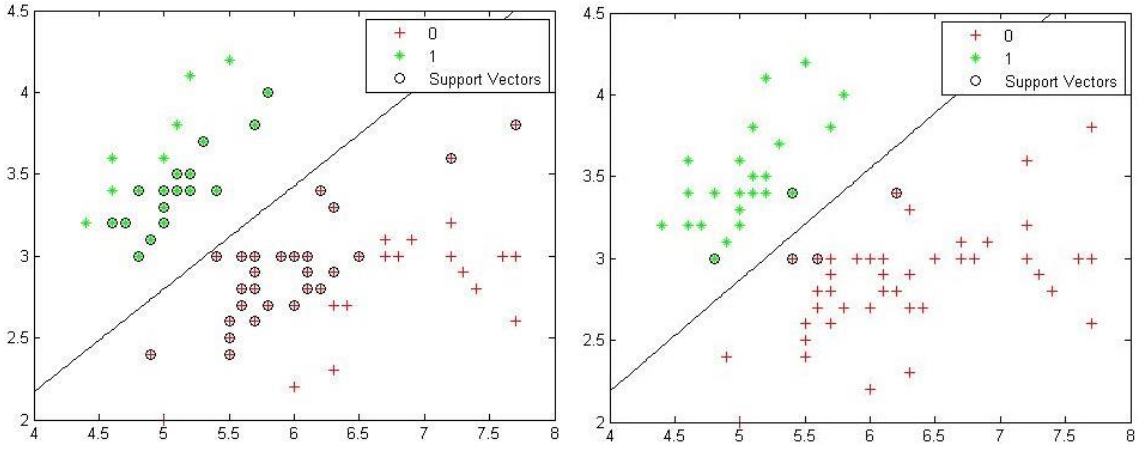

Figura 5-Kernel linear com C=0.1 á esquerda e C=10 á direita.

O *kernel* também tem um efeito determinante na fronteira de decisão (hiperplano). A largura do *kernel* Gaussiano e o grau do *kernel* polinomial afetam a flexibilidade do classificador. O grau do polinómio mais baixo é o *kernel* linear e com o aumento do grau do polinómio aumenta a curvatura da linha de fronteira (Ben-Hur & Weston, 2010), fig.6.

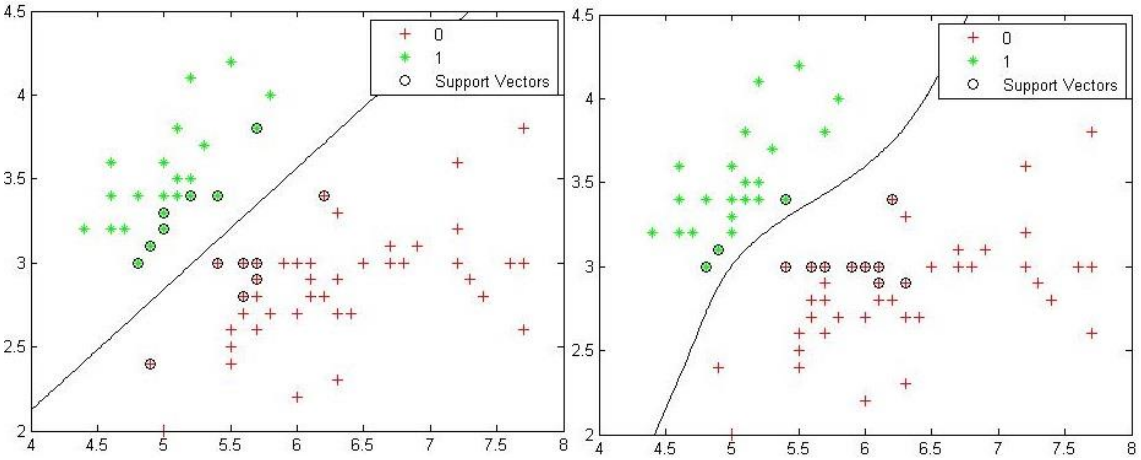

Figura 6-Kernel polinomial de ordem 1 á esquerda e ordem 5 á direita.

No caso do *kernel* Gaussiano um valor de gamma baixo torna a fronteira de decisão quase linear, á medida que o valor de gamma aumenta a flexibilidade desta fronteira também aumenta, valores altos de gamma levam a um *overfitting* (sobre ajustamento) dos dados (Ben-Hur & Weston, 2010), fig.7.

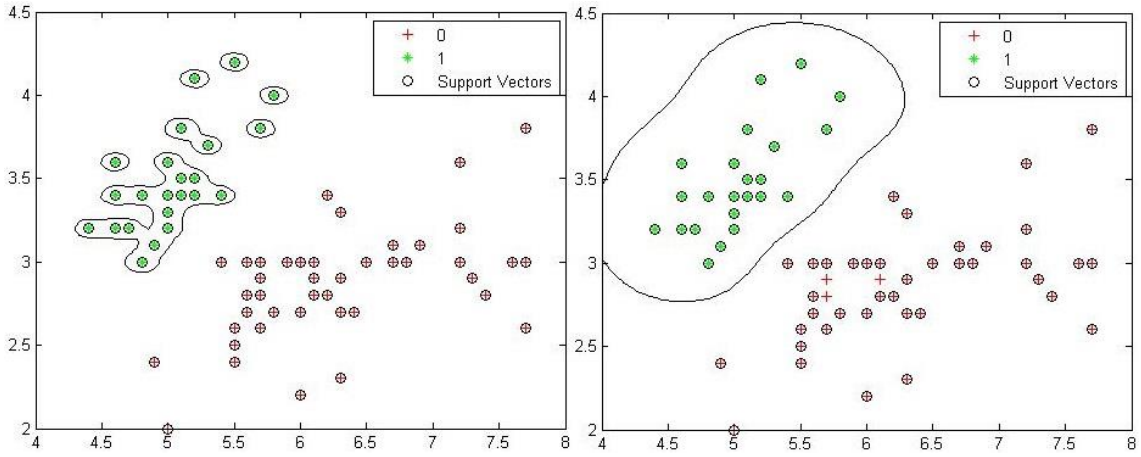

Figura 7-Kernel Gaussiano com sigma 0.1 á esquerda e 0.5 á direita.

Não existe uma regra para a escolha do *kernel*, portanto devem ser testados vários. Começando pelo linear e experimentando um não linear a ver se a performance melhora. A flexibilidade do *kernel* Gaussiano e polinomial normalmente leva a *overfitting* para conjuntos de dados grandes com baixo número de exemplos (Ben-Hur & Weston, 2010).

4.3. SELEÇÃO DE PARÂMETROS

A escolha de variáveis de entrada é uma consideração fundamental na identificação da forma funcional ótima dos modelos estatísticos. A tarefa de seleção de variáveis é comum ao desenvolvimento de todos os modelos estatísticos e é em grande parte dependente da descoberta de relações nos conjunto de dados disponíveis para identificar preditores adequados (May et al, 2011). Pretende-se explicar os dados da forma mais simples eliminando as variáveis redundantes. No caso da análise de regressão, isto implica que o modelo mais pequeno que se ajusta aos dados é o melhor. Variáveis desnecessárias irão acrescentar ruido á estimativa de outras quantidades em que estamos interessados. Tenta-se evitar a colinearidade que é causada pelo facto de ter muitas variáveis a tentar fazer o mesmo trabalho. Podemos poupar tempo e dinheiro reduzindo a dimensão do problema, tornando o sistema mais eficiente do ponto de vista computacional (Guyon & Elisseeff, 2003).

Em seguida são explicados os métodos de seleção de variáveis e redução de dimensão utilizados neste trabalho. O primeiro trata-se de um método rudimentar, em que é usado agrupamento por hierarquia (*hierarchical clustering*) com base no coeficiente de correlação e distância euclidiana. O segundo é a regressão linear passo a passo usando a função do Matlab *stepwisefit*. O terceiro é a análise das componentes principais, também conhecido como PCA. Por fim, foi ainda tentado um quarto método designado de *forward selection*,

computacionalmente muito exigente e estatisticamente pouco significativo pelo que não foram registados os resultados.

4.3.1. MÉTODO 1- AGRUPAMENTO POR HIERARQUIA, COEFICIENTE DE CORRELAÇÃO E DISTÂNCIA EUCLIDIANA

No método 1 foi usada uma técnica de *hierarchical clustering*, que traduzido á letra dá algo como agrupar por hierarquia. A ideia base do uso deste método foi a de fazer grupos com os parâmetros que estavam mais correlacionados entre si, usando a função do Matlab *corrcoef*. Existe um valor que pode ser ajustado, que determina a distância a partir da qual um elemento deve pertencer ao grupo ou não. No caso esse valor, designado de *cut off*, foi ajustado para 0.5. De seguida era selecionado apenas um parâmetro de cada grupo segundo a distância euclidiana. Aquele que tivesse maior distância euclidiana era selecionado. A distância euclidiana foi calculada para todos os parâmetros e entre os conjuntos a serem testados, patológico ou saudável.

4.3.2. MÉTODO 2-REGRESSÃO LINEAR PASSO A PASSO

Foram desenvolvidos métodos de seleção de variáveis que permitem encontrar bons subconjuntos de variáveis (modelos) usando menos recursos computacionais do que todos os outros tipos de regressão. Estes métodos são conhecidos como métodos de regressão passo a passo (Rawlings et al, 1998). A regressão passo a passo começa por escolher uma equação contendo uma única variável com mais significado. Depois vai adicionando variáveis, uma de cada vez, enquanto as adições trouxerem melhorias ao modelo. A ordem da adição é garantida pelo teste F que permite selecionar qual a próxima variável a entrar. O valor mais alto do teste F é comparado com o valor de teste F definido por nós ou por defeito. Após a variável entrar no modelo, a equação é examinada para ver se existe alguma variável que possa ser excluída (Draper & Smith, 1998).

No método 2 é feita uma análise por regressão linear múltipla usando uma função do Matlab chamada de *stepwisefit*. Trata-se de um método sistemático para a adição e remoção de termos de um modelo multilinear com base na sua significância estatística. O método começa com um modelo inicial e em seguida, compara o poder explicativo dos modelos, maiores ou menores, incrementados. Em cada etapa, o valor p de uma estatística do teste F é calculado para testar modelos com e sem um termo potencial. Se um termo não está neste momento no modelo, a hipótese nula é que o termo teria um coeficiente de zero se adicionado ao modelo.

Se houver evidência suficiente para rejeitar a hipótese nula, o termo é adicionado ao modelo. Por outro lado, se um termo está atualmente no modelo, a hipótese nula é que o termo tem um coeficiente de zero. Se não há provas suficientes para rejeitar a hipótese nula, o termo é removida do modelo. O p valor máximo para um termo ser adicionado foi fixado em 0.05 e o p valor mínimo para um termo ser removido ficou em 0.10. No final é devolvido um modelo com os termos/variáveis que serão usadas na rede neuronal (Mathworks, Support, 2016), (Rodríguez, 2010).

4.3.3. MÉTODO 3 - PCA

No método 3 é usada uma técnica de redução de dimensão chamada de Principal Components Analysis (PCA). É uma técnica estatística que usa conceitos matemáticos como o desvio padrão, a covariância e os valores e vetores próprios. Tem aplicações nos campos do reconhecimento facial e compressão de imagem, sendo uma técnica comum para encontrar padrões em dados de grande dimensão. Identifica padrões nos dados e expressa-os de forma a realçar as suas semelhanças e diferenças. Primeiramente é subtraída a média para cada dimensão dos dados, isto produz um conjunto cuja média é zero, designados de dados ajustados. Seguidamente são calculados os valores e vetores próprios a partir da matriz de covariância. Depois tem que se decidir quantas componentes vão ser selecionadas. Como foi usada a função do Matlab *princomp* para calcular as componentes principais e esta devolve os valores próprios de forma ordenada é só calcular a percentagem cumulativa destes valores. São então selecionadas os primeiros vetores próprios correspondentes a 90 ou 95 % da percentagem cumulativa. Isto significa que esses primeiros vetores próprios explicam 90 a 95 % dos dados. Por fim são multiplicados os dados ajustados pela inversa da matriz dos vetores próprios selecionados (Smith, 2002). Para ter resultados mais próximos do real a média dos dados foi calculada apenas no conjunto de treino e subtraída aos conjuntos de validação e de teste separadamente.

4.3.4. FORWARD SELECTION

Foi também criado um algoritmo para fazer a seleção de variáveis designado de *forward selection*. *Forward selection* é uma estratégia de pesquisa incremental linear que seleciona possíveis variáveis candidatas uma de cada vez. O método começa treinando a rede com modelos de variável única selecionando a variável que maximiza a performance. O processo continua adicionando uma variável de cada vez ao modelo e selecionando aquela que melhor

performance acrescenta ao modelo anterior. A seleção termina quando determinada performance é alcançada ou quando a variável acrescentada falha no aumento da performance do modelo atual. Este método apresenta algumas debilidades, como não testar possíveis combinações que seriam melhores, uma vez que é feito de forma linear, mas o principal problema é a morosidade do processo (May et al, 2011). Este método foi testado mas devido á lentidão do processo e a falta de rigor estatístico foi posto de parte.

CAPÍTULO V

5. DESENVOLVIMENTO

5.1. EXTRAÇÃO DE PARÂMETROS

Neste trabalho foram estudados vários parâmetros e a sua capacidade de distinção entre saudável e patológico. Foram extraídos pelo menos dois conjuntos de parâmetros distintos, parâmetros de análise de longo termo e de curto termo. O primeiro conjunto de parâmetros envolvem quatro medidas de Jitter, Shimmer e HNR para três tons e três vogais diferentes. O segundo conjunto de parâmetros correspondem aos coeficientes cepstrais na frequência mel, frequências e larguras de banda dos três primeiros formantes, frequência fundamental, energia, momentos espectrais de ordem zero, um, dois, três e curtose e a potência.

5.1.1. ALGORITMO

O algoritmo desenvolvido por Gonçalves (Teixeira & Gonçalves, 2016) permite a extração de 9 parâmetros. O jitter absoluto, jitter relativo, jitter rap, jitter ppq5, shimmer absoluto, shimmer relativo, shimmer apq3, shimmer apq5 e HNR. Este algoritmo foi usado para extrair estes 9 parâmetros para três vogais e três tons diferentes a partir de sinais disponíveis na base de dados SVD. As vogais disponíveis são /a/, /i/, /u/ e os tons baixo, normal e alto.

Foi então criado um programa em código Matlab que permitia a extração de todos parâmetros de todos os sinais disponíveis na base de dados. Para tal foi necessário proceder á catalogação de todas as doenças atribuindo um número de 1 a 150, o número de doenças disponível, a partir da sua ordem alfabética. A matriz (p) devolvida pelo algoritmo referido contém por coluna os 9 parâmetros, o código numérico indicativo da vogal, tom e sexo do paciente. Foi criada também uma matriz (Plabel) que contém o rótulo que indicava se era controlo ou patológico, qual a patologia, vogal e tom a que pertenciam aqueles 9 parâmetros. Por exemplo Can (Controlo vogal /a/ tom normal) ou P40an (Disfonia vogal /a/ tom normal). Este rótulo permite uma pesquisa facilitada e extração de dados da matriz principal para uso futuro em testes com classificadores.

5.1.2. CONJUNTO DE PARÂMETROS ALTERNATIVO

Foram criados uma série de scripts em código Matlab que permitiam a extração de 12 coeficientes cepstraais na frequência mel, frequência e largura de banda dos 3 primeiros formantes, frequência fundamental, energia, momentos espectrais de ordem 0, 1, 2, 3 e curtose e a potência *root mean square*. Todos os scripts têm em comum: o tamanho da janela de análise e os intervalos, definidos para 20 (ms) e 10 (ms) respetivamente. Existe uma sobreposição de janelas de 50% (*overlapping*); Reamostragem do sinal baixando a frequência de amostragem para 16 kHz; Seleção de uma janela de sinal de 500 (ms) com base na posição da energia máxima, usando a energia deslizante. Isto faz com que seja selecionada apenas a parte onde existe sinal evitando zonas de silêncio; Cálculo da média ao longo das várias *frames* para cada um dos parâmetros mencionados.

A extração dos coeficientes cepstrais a frequência mel era feita recorrendo á função disponível em: https://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab (Mathworks, Community, File Exchange, 2016). A função devolve os coeficientes cepstrais na frequência mel (MFCC). O sinal é primeiro pré-enfatizado usando um filtro FIR de primeira ordem através de um coeficiente de pré-enfase fornecido pelo utilizador. O sinal pré-enfatizado é sujeito a uma análise de curto termo usando a transformada de Fourier e tamanho de janela e intervalos especificados pelo utilizador. Em seguida é calculada a potência espectral e aplicado o banco de filtros triangulares uniformemente espaçados entre a frequência mínima e máxima na escala mel. Como ultimo passo é aplicado um filtro sinusoidal. Os parâmetros de entrada usados foram os seguintes: Pré-enfase (0,97), gama de frequências a considerar na análise [300,3700], número de canais do banco de filtros (20) e filtro sinusoidal (22).

O algoritmo utilizado neste trabalho para o cálculo das frequências formantes e respetivas larguras de banda é baseado no exemplo presente na documentação do Matlab (Mathworks, Support, 2016) (http://www.mathworks.com/help/signal/ug/formant-estimation-with-lpc-coefficients.html). Em seguida serão descritos os passos para o cálculo destes parâmetros. Em primeiro lugar são aplicadas duas técnicas comuns no processamento de sinais de fala. É aplicada uma janela de Hamming e um filtro de pré-enfase. O filtro de pré-enfase é um filtro passa alto só com pólos. Em seguida são determinados os coeficientes de predição linear usando a função do Matlab *lpc*, na qual é necessário especificar a ordem do LPC. Existe uma regra geral que diz que a ordem do LPC deve ser duas vezes o número de formantes esperados

mais dois (Mathworks, Support, 2016). Contudo, a ordem terá que ser adaptada á frequência de amostragem e ao propósito. Em seguida é necessário achar as raízes do polinómio devolvido pela função do Matlab *lpc*. As raízes vão aparecer na forma de pares complexos conjugados. Devem ser retidas apenas as raízes com um sinal na parte imaginaria e calculados as fases dessas raízes. Converter as frequências angulares em radianos/segundo, representadas pelos ângulos, para Hz e calcular as larguras de banda dos formantes. As larguras de banda dos formantes são representadas pela distância dos zeros do polinómio ao círculo unitário. É usado o critério de que as frequências dos formantes devem ser superiores a 90 Hz com larguras de banda inferiores a 400 Hz para determinar os formantes (Mathworks, Support, 2016). A ordem do *lpc* usada foi 18. O envelope espectral pode ser definido como uma curva que vai ligar os picos do espectro. Estes picos vão definir os formantes. Se a ordem do *lpc* for demasiado baixa aquilo que deveriam ser dois formantes passa a ser apenas um. Por outro lado se a ordem for demasiado alta aquilo que deveria ser apenas um formante passa a ser dois. Como tal é necessário encontrar um equilíbrio. Tendo por base valores referência de uma análise feita com o Praat foi possível ajustar a ordem para o valor acima descrito. Tentando assim evitar os problemas associados á ordem do *lpc* descritos.

O script que faz a extração da frequência fundamental (F0) recorre ao método da auto correlação. Como está descrito no capítulo 3 (parâmetros/F0) este método pressupõe que se ache o máximo da função de autocorrelação. É também aplicada uma janela de hamming do mesmo tamanho da frame em análise.

O script que faz o cálculo da energia do sinal aplica a eq. 25.Também aqui é aplicada uma janela de hamming do mesmo tamanho da frame em análise.

No que respeita á extração dos momentos espectrais de ordem 0, 1, 2, 3 e curtose. Primeiro é usada a função *periodogram* do Matlab que devolve a potência espectral em função da frequência. Em seguida é aplicado o log10 para passar para uma escala logarítmica e termos a potência em dB (Figura 8). Depois são calculados os momentos e curtose usando as eqs. 18-22 mas apenas para os valores de potência até aos 4,5 kHz.

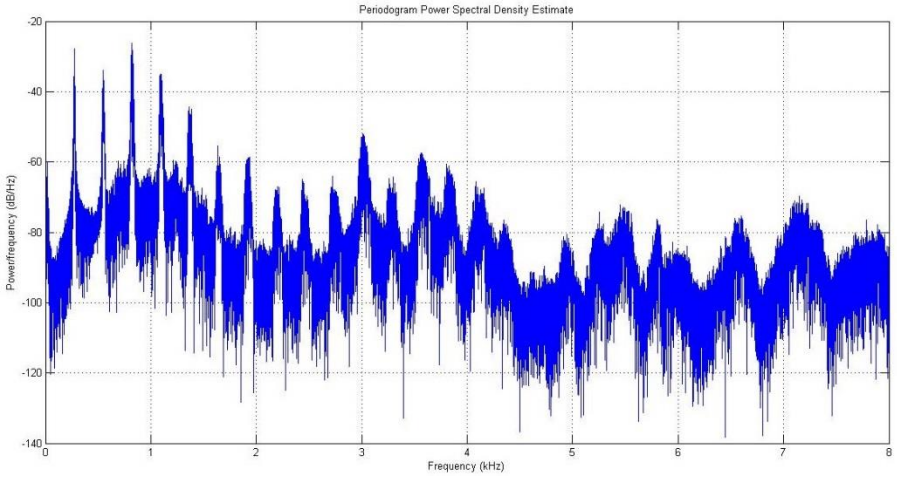

Figura 8-Periodograma da Potência Espectral.

No script que calcula a potência é extraída a potência *root mean square* de acordo com a eq.29.

Foram feitos alguns testes com estes parâmetros usando uma técnica, que ao contrário da média das *frames*, usava todas as *frames* para treinar o classificador. Depois era feito um pós-processamento para apurar a precisão que pressupunha que se X% (*threshold)* das *frames* de um sinal forem assinalados como pertencentes a uma classe então todo o sinal é considerado dessa classe. Para esta experiência eram excluídos 100 ms do início e 100 ms do fim do sinal garantindo assim, na maior parte dos casos, maior número de *frames* que os 500 ms descritos anteriormente.

5.1.3. PARÂMETROS EXTRAÍDOS COM O PRAAT

O Praat é um programa de computador que permite analisar, sintetizar e manipular sons de voz. Foi desenvolvido por Paul Boersma e David Weenick em 1992 no Instituto de Ciência e Fonética da Universidade de Amesterdão (Boersma et al, 2001). É uma aplicação gratuita e permite a análise espectral (espectrografia), análise dos parâmetros de F0, jitter, shimmer e HNR, análise dos formantes e análise da intensidade do sinal (Lopes, 2008). Este *software* permite ainda a criação de scripts para execução e análise de grandes quantidades de dados de forma automática (Boersma et al, 2001). Na figura 9 podemos ver a interface gráfica deste programa.

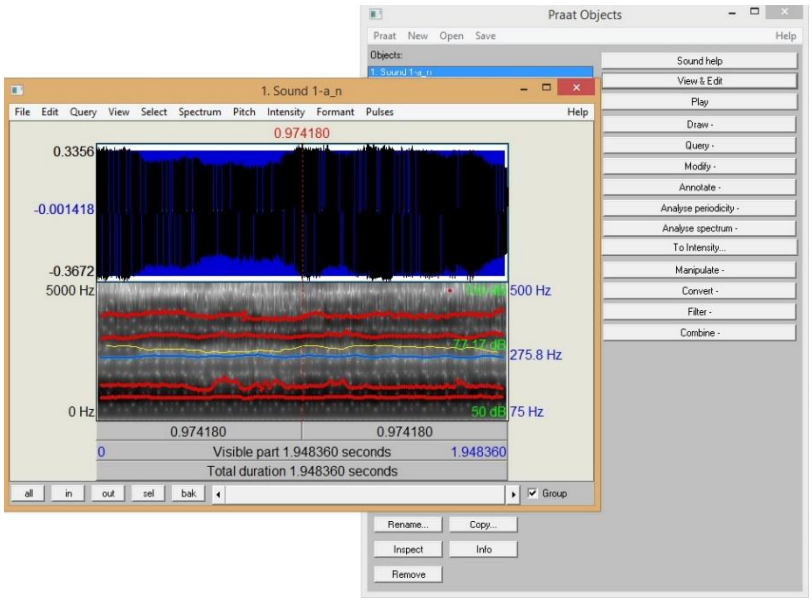

Figura 9-Interface gráfica do programa Praat.

O *software* Praat permite a criação de scripts para execução e análise de grandes quantidades de dados de forma automática. Foram então criados três scripts diferentes para execução com o Praat. Para todos eles é necessário fornecer o diretório de localização e os dados são apresentados no ecrã.

O primeiro permite a extracção de 9 parâmetros. O jitter absoluto, jitter relativo, jitter rap, jitter ppq5, shimmer absoluto, shimmer relativo, shimmer apq3, shimmer apq5 e HNR. Este script contém no seu código as diretrizes necessárias á extração destes parâmetros mediante fornecimento do diretório onde se encontram os sinais. Primeiro é feita uma análise daF0 fornecendo como parâmetros o intervalo de análise (*Time step*) em segundos, o F0 mínimo (F0 *floor*), que determina o tamanho da janela de análise, e máximo (F0 *ceiling*) em Hertz (Paul Boersma, Manual Praat, 2003). O Time step foi de 0.0 o que faz com que segundo o Praat seja utilizado um intervalo de 0,75/(F0 *floor*). A Frequência fundamental *floor* e *F0 ceiling* foram 75 e 600 Hz respetivamente. Em seguida é feita a extração de um relatório extensivo sobre vários parâmetros do sinal (*voice report*) do qual são extraídos apenas os parâmetros acima indicados. Os dados são apresentados no ecrã e posteriormente é criada uma matriz no Matlab para futura aplicação em classificadores inteligentes.

O segundo permite a extração de 12 coeficientes de cepstrais na frequência mel. Este script foi adaptado a partir de um script desenvolvido por Jeff Mielk (disponível em: http://phon.chass.ncsu.edu/manual/wav2mfcc.praat). A análise dá-se em dois passos. Primeiro

passo, é feita uma análise espectral na escala mel. Segundo passo, os valores do espectrograma são convertidos em coeficientes cepstrais na frequência mel (Paul Boersma, Manual Praat, 2003). Os parâmetros do banco de filtros usados foram: 100 mel para a posição do primeiro filtro, 100 mel para a distância entre filtros e 0.0 mel para a frequência máxima. O tamanho da janela de análise foi de 15 ms com intervalos (Time step) de 5 ms. Estes valores são os valores padrão do programa.

O terceiro permite a extração da frequência e largura de banda dos 3 primeiros formantes (Hz), F0/*Frequência fundamental* (Hz), energia (Pa2.seg), potência (Pa2), intensidade (dB), momentos espectrais de ordem 1, 2 e 3 e curtose (Hz). As frequências e larguras de banda dos formantes foram obtidas no centro do sinal. O *Frequência fundamental* corresponde ao *Frequência fundamental* médio do sinal. A energia, potência e intensidade foi calculada para o sinal todo. O momento de ordem 1, também conhecido como centro de gravidade, mede o quão alto são as frequências em média no espectro. O momento de ordem 2, também conhecido como desvio padrão, mede o desvio das frequências no espectro em relação ao centro de gravidade. O momento de ordem 3 é uma medida de assimetria que mede a diferença entre a forma do espectro abaixo do centro de gravidade e a forma do espectro acima da frequência média. A curtose mede a diferença entre a forma do espectro, relacionada com o centro de gravidade, e a forma Gaussiana (Paul Boersma, Manual Praat, 2003).

5.2. IMPLEMENTAÇÃO DA RNA EM MATLAB

A implementação da Rede Neuronal Artificial (RNA) foi feita em código Matlab. Este programa dispõe de uma série de funções que permite criar, treinar e simular a rede. Existe ainda uma *toolbox* para uso de redes neuronais muito intuitiva e sem necessidade de recurso a quase nenhuma linha de código. Contudo, optou-se pela escrita de um código. Nele constam o carregamento dos dados e criação das matrizes necessárias para alimentar a rede. A matriz p com os exemplos de treino (*inputs*) e a matriz t com os alvos (*targets*). Na matriz t constavam zeros atribuídos aos pacientes de controlo ou saudáveis e uns atribuídos aos pacientes patológicos. Foi também feita uma divisão dos dados em três subconjuntos.

O primeiro subconjunto é o de treino, o qual é usado para calcular o gradiente e atualizar os pesos e desvios. O segundo subconjunto é o de validação. O erro no subconjunto de validação é monitorizado durante o processo de treino. O erro associado á validação normalmente desce durante a fase inicial de treino, assim como o erro associado ao conjunto de treino. Contudo, quando a rede começa a sobre ajustar os dados (*overfit*), o erro no conjunto de validação

aumenta. São então guardados os pesos e desvios associados ao erro mínimo do conjunto de validação. O terceiro subconjunto é o de teste, este não é "visto" durante a fase de treino e serve para apurar o poder de predição da rede após treino (Mathworks, Support, http://www.mathworks.com/help/nnet/ug/divide-data-for-optimal-neural-network-training.html?searchHighlight=neural%20network%20data%20division).

Optou-se por usar 70% dos dados para treino, 15% para validação e 15% para teste.

Foi criado um ciclo que permitia 20 repetições da inicialização da rede, treino e simulação. Após este ciclo era guardada a rede que apresentava melhor precisão com base no conjunto de validação (netfinal). A criação deste ciclo deve-se ao facto de a inicialização dos pesos da rede ser feita de forma aleatória o que faz com que se obtenham resultados diferentes cada vez que é executado todo o processo. A precisão era calculada com base na equação 40 e como na saída da rede nem sempre temos exatamente zeros e uns foi necessário proceder a algum processamento. Primeiro os valores são arredondados e em seguida os valores ≤ 0 passam a 0, os valores ≥ 1 passam a 1. E assim obtemos só zeros e uns. Ou seja <0,5 passa a 0 e >=0,5 passa a 1. Para calcular o valor de r (coeficiente de correlação) foi usada a função do Matlab *corrcoef* que devolve uma matriz com os coeficientes de correlação, bastando depois obter a triangular superior uma vez que esta é espelhada sobre a diagonal principal.

$$\text{Precisão} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FN} + \text{VN} + \text{FP}} \tag{40}$$

Além da precisão, dada pela eq. 40, foram ainda usadas a sensibilidade e especificidade para avaliar os resultados calculadas de acordo com o que esta presente na tabela 2.

Tabela 2-Matriz de confusão usada na análise dos resultados.

		Resultados da classificação	
		Saudável	**Patológico**
Diagnóstico	**Saudável**	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Patológico	Falso Negativo (FN)	Verdadeiro Negativo (VN)
		Sensibilidade = VP/(VP+FN)	**Especificidade**=VN/(VN+FP)

Para o caso em que são usados 2 neurónios na camada de saída a codificação atribuída é $\begin{matrix}0\\1\end{matrix}$ para o controlo e $\begin{matrix}1\\0\end{matrix}$ para o patológico. No pós-processamento da saída é calculado o máximo e registada a posição do máximo. Se a posição do máximo for a segunda é controlo, por outro lado, se estiver na primeira posição é patológico. Outro tipo de pós processamento da saída usado foi o valor mais próximo de 1. É calculado o valor mais próximo de 1 e registada a posição, a partir daí é igual ao pós processamento com o máximo.

5.3. IMPLEMENTAÇÃO DA MVS EM MATLAB

A implementação da Máquina de Vetor de Suporte foi feita em código Matlab usando duas funções principais a *svmtrain* e a *svmclassify*. A primeira permite treinar a MVS e a segunda serve para apurar o poder de predição do classificador após treino. Com a MVS é necessário dividir a matriz de entrada em dois subconjuntos, o de treino e o de teste. A percentagem usada para treino foi de 85% e para teste 15%.

Foram criados alguns ciclos para gerar todas as combinações de parâmetros de entrada da função *svmtrain* e guardar numa matriz o conteúdo dos testes realizados. As combinações referem-se aos diferentes tipos de *Kernel*, parâmetros associados a esses *Kernels* e diferentes métodos que permitem encontrar o hiperplano de separação. Na matriz constavam os resultados dos testes indicando qual o *kernel* usado, os parâmetros de entrada, método, precisão do conjunto de teste, sensibilidade e especificidade. O cálculo da precisão, sensibilidade e especificidade estão de acordo com o que é indicado na Tabela 1 apesar do uso da função *classperf* que permite avaliar a performance do classificador. Os métodos são o *Quadratic Programming* (QP), *Sequential Minimal Optimization* (SMO) e *Least Squares* (LS).

CAPÍTULO VI

6. RESULTADOS E DISCUSSÃO

Nesta secção vão ser relatados os resultados de algumas experiencias feitas assim como a discussão das mesmas. Foram analisadas as respostas dos classificadores a dois conjuntos de parâmetros diferentes extraídos por diferentes algoritmos. O conjunto de parâmetros 1 onde constam: HNR, quatro medidas de jitter e quatro medidas de shimmer. E o conjunto de parâmetros 2 onde constam: 12 coeficientes cepstrais na frequência mel (MFCC), frequências e larguras de banda dos três primeiros formantes, energia, potencia, momentos espectrais de ordem zero, um, dois, três e curtose. O conjunto de parâmetros 1 foi extraído pelo algoritmo desenvolvido por Gonçalves (Teixeira & Gonçalves, 2016) e também pelo Praat. O conjunto de parâmetros 2 foi extraído por um algoritmo desenvolvido nesta tese e também pelo Praat. Foram aplicadas algumas técnicas para seleção de parâmetros. Ao longo desta tese foi usado o termo "modelos" como se referindo aos parâmetros de entrada encontrados pela aplicação destas técnicas. Estes modelos ou parâmetros de entrada podem ser vistos na tabela 11.

6.1. DESCRIÇÃO DAS EXPERIÊNCIAS COM O ALGORITMO

O algoritmo desenvolvido por Gonçalves (Teixeira & Gonçalves, 2016) foi usado para extrair os seguintes parâmetros: jitter absoluto, jitter relativo, jitter ppq5, jitter rap, shimmer absoluto, shimmer relativo, shimmer apq3, shimmer apq5 e HNR. Para as vogais /a/, /i/ e /u/ e tons baixo, normal e alto. Os 9 parâmetros extraídos para as três vogais e três tons diferentes foram então organizados num vetor coluna com 81 variáveis por paciente.

Foram treinados dois classificadores diferentes, RNA e MVS, utilizando como parâmetros de entrada todos os parâmetros e a combinação de parâmetros determinada com o método 1, 2 e PCA. Os dois primeiros métodos referem-se a métodos de seleção de variáveis e o PCA a uma técnica de redução da dimensão.

Foi testada a capacidade de predição dos classificadores para duas doenças diferentes, disfonia (P40) e paralisia das cordas vocais (P136). Foram usadas estas duas patologias porque eram as que apresentavam maior número de sujeitos na base de dados. Os classificadores foram treinados com o mesmo número de exemplos positivos (Controlo) e

negativos (Patológico). Foi feita a separação por género, ficando assim com controlo feminino e masculino (CF e CM), disfonia feminino e masculino (P40F e P40M) e paralisia feminino e masculino (P136F e P136M).

Foram testadas várias topologias e combinações tanto na RNA como MVS. Devido á inicialização aleatória dos pesos, na RNA foram usados 20 ciclos de treino e guardada a rede com melhor precisão com base no conjunto de validação. A seguir são apresentados os melhores resultados.

Numa primeira experiencia foram usados os parâmetros supracitados para varias vogais e tons e numa segunda experiencia foram usados esses mesmos parâmetros apenas na vogal /a/ no tom normal. O objetivo era aferir se o uso de várias vogais e vários tons era uma mais-valia em relação a apenas uma vogal e um tom.

6.1.1. RESULTADOS

Nas tabelas 3, 4, 5 e 6 temos os melhores resultados para a RNA, com os 9 parâmetros vezes 3 vogais e 3 tons (todos os parâmetros) e com os modelos achados com a aplicação dos métodos 1, 2 e 3. A aplicação dos métodos/técnicas de seleção de variáveis e redução da dimensão prendem-se com o facto de o vetor de entrada ser de grande dimensão e com a possibilidade de achar as variáveis que melhorem a capacidade de predição da rede.

Para a classificação entre saudável e patológico foram usados 4 grupos diferentes, disfonia feminino (P40F, tabela 3), difonia masculino (P40M, tabela 4), paralisia das cordas vocais feminino (P136F, tabela 5) e paralisia das cordas vocais masculino (P136M, tabela 6).

Nas tabelas 3, 4, 5 e 6 temos informação sobre os parâmetros usados na entrada da rede (Entrada), a arquitetura usada com o número de neurónios da entrada, camada escondida e saída (Arq. [E,CE,S]), função de transferência da camada escondida (FTCE), função de transferência da saída (FTS), função de treino (FT), valor de R e Precisão do conjunto de treino mais validação (R-P2 e Prec.-P2) e valor de R e Precisão do conjunto de teste (R-T e Prec.-T).

Analisando a tabela 3 podemos ver que existe uma melhoria da precisão do conjunto de teste de 83.3 para 100% devido ao uso do Modelo 4 achado com o método 2. O método 3 (PCA) consegue igualar o resultado obtido com o uso de todos os parâmetros, demonstrando que é

possível reduzir a dimensão do problema sem perdas de informação. Foi ainda testada a possibilidade de serem usados 2 neurónios na camada de saída com diferentes processamentos da saída, com o máximo e com o valor mais próximo de 1. Para o caso em que são usados 2 neurónios na camada de saída a codificação atribuída é $\begin{matrix}0\\1\end{matrix}$ para o controlo e $\begin{matrix}1\\0\end{matrix}$ para o patológico. Como podemos ver na tabela 3 o uso de 2 neurónios na camada de saída, quer com um processamento quer com o outro, não apresentou melhores resultados do que com o uso de 1 só.

Tabela 3-Resultados da RNA para a disfonia feminino (CFvsP40F).

Entrada	Arq. [E,CE,S]	FTCE	FTS	FT	R-P2	Prec.-P2 [%]	R-T	Prec.-T [%]
Todos param.	[81,20,1]	tansig	purelin	trainlm	0,85	92,7	0,67	83,3
Método 1-Modelo 6	[17,10,1]	tansig	purelin	trainlm	0,66	81,7	0,71	83,3
Método 2-Modelo 4	[6,15,1]	tansig	purelin	trainscg	0,69	84,1	1	100
Método 3-PCA	[7,15,1]	tansig	purelin	trainlm	0,53	75,6	0,67	83,3
Todos param.	[81,20,2]*	tansig	purelin	trainlm	-	62,2	-	66,7
Todos param.	[81,20,2]**	tansig	purelin	trainlm	-	92,7	-	66,7

*pós processamento da saída com o máximo.

**pós processamento da saída com o valor mais próximo de 1.

Pela análise da tabela 4 podemos ver que, á semelhança da tabela 3, houve uma melhoria dos resultados usando os métodos de seleção de variáveis e redução da dimensão. A precisão do conjunto de teste passou de 70 para 90%. Mais uma vez o método 2 alcança melhores resultados do que o método 1 mas desta vez o PCA também consegue igualar esses resultados.

Tabela 4-Resultados da RNA para a disfonia masculino (CMvsP40M).

Entrada	Arq. [E,CE,S]	FTCE	FTS	FT	R-P2	Prec.-P2 [%]	R-T	Prec.-T [%]
Todos param.	[81,10,1]	logsig	purelin	trainlm	0,04	51,7	0,50	70,0
Método 1-Modelo 8	[14,20,1]	logsig	purelin	trainlm	0,33	65,5	0,41	70,0
Método 2-Modelo 5	[2,15,1]	logsig	purelin	trainlm	0,39	67,2	0,82	90,0
Método 3-PCA	[7,10,1]	tansig	purelin	trainscg	0,32	65,5	0,82	90,0

A tabela 5 apresenta-nos os melhores resultados para o grupo da paralisia das cordas vocais feminino. A sua observação permite-nos aperceber de que não ocorreram melhorias pela

aplicação dos métodos de seleção de variáveis. Contudo, as perdas não são assim tão significativas contando com valores da ordem dos 2,6% de melhoria em relação ao método 1 e PCA.

Tabela 5-Resultados da RNA para a paralisia das cordas vocais feminino (CFvsP136F).

Entrada	Arq. [E,CE,S]	FTCE	FTS	FT	R-P2	Prec.-P2 [%]	R-T	Prec.-T [%]
Todos param.	[81,10,1]	tansig	purelin	trainlm	0,694	84,5	0,527	76,3
Método 1-Modelo 9	[7,15,1]	tansig	purelin	trainlm	0,566	76,9	0,484	73,7
Método 2-Modelo 10	[6,20,1]	tansig	purelin	trainlm	0,613	79,4	0,436	71,1
Método 3-PCA	[7,10,1]	tansig	purelin	trainlm	0,549	77,4	0,476	73,7

Na tabela 6 temos os resultados da RNA para a paralisia das cordas vocais masculino. Pela análise desta verificamos que ocorreram melhorias, de 77,3 para 81,8%, pela aplicação dos métodos 2 e 3 (PCA).

Tabela 6-Resultados da RNA para a paralisia das cordas vocais masculino (CMvsP136M).

Entrada	Arq. [E,CE,S]	FTCE	FTS	FT	R-P2	Prec.-P2 [%]	R-T	Prec.-T [%]
Todos param.	[81,15,1]	logsig	purelin	trainlm	0,619	80,4	0,567	77,3
Método 1-Modelo 11	[10,15,1]	logsig	purelin	trainlm	0,651	81,2	0,462	72,7
Método 2-Modelo 12	[3,25,1]	tansig	purelin	trainscg	0,583	78,9	0,647	81,8
Método 3-PCA	[7,15,1]	logsig	purelin	trainlm	0,566	78,3	0,636	81,8

Como em grande parte dos casos o método 2 e PCA registam melhores resultados do que o método 1, e por este ser um pouco trabalhoso e quase manual, foi posto de lado e a partir desta fase foram usados apenas os métodos 2 e 3 (PCA).

Nas tabelas 7, 8, 9 e 10 podemos ver os melhores resultados alcançados para a MVS. Estas tabelas apresentam informação sobre o tipo de *kernel* utilizado (kernel), os parâmetros associados a esse *kernel* (Parâm.), o método de treino (Mét.), a precisão (Prec.-T), sensibilidade (Sens.) e especificidade (Espec.) do conjunto de teste.

Na tabela 7 podemos visualizar os melhores resultados alcançados pela MVS para o grupo da disfonia feminino. A aplicação do método 2 produziu um modelo que foi capaz de melhorar os resultados de 83.3 para 100%. O método 3 (PCA) conseguiu igualar o resultado obtido

com todos os parâmetros, demonstrando que é possível reduzir a dimensão do problema sem perdas.

Tabela 7-Resultados da MVS para a disfonia feminino (CFvsP40F).

Entrada	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
Todos param.	linear	C=0,1	SMO	91,7	100	83,3
Método 2-Modelo 4	linear	C=0,1	QP	100	100	100
Método 3-PCA	Gauss.	S=2, C=10	QP	83,3	83,3	83,3

Na tabela 8 podemos observar os melhores resultados da MVS para a disfonia masculino. Mais uma vez verifica-se que a aplicação do método 2 consegue melhorar significativamente os resultados passando de 75 para 100%. O PCA volta a alcançar o mesmo resultado do que quando se usa todos os parâmetros na entrada.

Tabela 8-Resultados da MVS para a disfonia masculino (CMvsP40M).

Entrada	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
Todos param.	linear	C=0,1	SMO	87,5	100	75
Método 2-Modelo 5	Gauss.	S=0,1, C=0,2	QP	100	100	100
Método 3-PCA	linear	C=1	SMO	87,5	100	75

A tabela 9 apresenta-nos os resultados da MVS para a paralisia das cordas vocais feminino. A sua análise permite-nos dizer que, á semelhança dos resultados RNA para o mesmo grupo, não há melhorias após aplicação dos métodos 2 e 3. Também os resultados apresentados ficam um pouco abaixo do pretendido.

Tabela 9-Resultados da MVS para a paralisia das cordas vocais feminino (CFvsP136F).

Entrada	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
Todos param.	Poli.	O=2, C=0,04	SMO	78,9	84,2	73,7
Método 2-Modelo 10	Poli.	O=2, C=10	QP	78,9	89,5	68,4
Método 3-PCA	Gauss.	S=1, C=0.1	QP	76,3	84,2	68,4

Os resultados contidos na tabela 10 pertencem á utilização da MVS com a paralisia das cordas vocais masculino. A aplicação do método 3 (PCA) permitiu uma melhoria de 70 para 80%, ficando o método 2 com valores inferiores a todos os outros, apenas 60%.

Tabela 10-Resultados da MVS para a paralisia das cordas vocais masculino (CMvsP136M).

Entrada	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
Todos param.	Gauss.	S=4, C=0,2	QP	75,0	80,0	70,0
Método 2-Modelo 12	linear	C=0,1	QP	75,0	90,0	60,0
Método 3-PCA	Poli.	O=4, C=1	QP	80,0	80,0	80,0

Na tabela 11 podemos observar o conjunto de parâmetros (modelos) selecionados pela aplicação dos métodos/técnicas de seleção de variáveis. Em relação ás várias medições de Jitter podemos afirmar que o Jitter absoluto é o que mais vezes é selecionado e em mais vogais e tons diferentes. Para as quatro medidas de Shimmer apresentadas o Shimmer relativo é o que mais poder de predição apresenta entre as quatro medidas possíveis, associado a várias vogais e tons. Por fim o HNR também é selecionado com frequência para várias vogais e tons. Estes três parâmetros parecem ser os mais relevantes podendo ser usados com grande parte das vogais e tons.

Tabela 11-Modelos encontrados com a aplicação das técnicas de seleção de variáveis.

Modelos / Parâmetros	6	4	8	5	9	10	11	12
Jitta	ul,ih,an,un,in,ah,al		in,an,il,ih,al,uh		al,un,ah,an,ul,il	an	un,il,ul,ih,an,ah	ah
Jitter				al				
Rap				ah				in
Ppq5						un		
ShdB								
Shim	il,ul,in,ih,an	in	ah,al,an,in			uh	un,ah,al	
Apq3		an,ah,uh				ah		
Apq5		in						
HNR	ih,il,ul,al	al	al,ul,ah,un		al	al,in		an

Numa segunda experiencia o Algoritmo desenvolvido por (Teixeira & Gonçalves, 2016) foi utilizado para extrair os mesmos parâmetros da experiencia anterior mas desta vez apenas para a vogal /a/ no tom normal. Estes parâmetros foram usados para treinar dois classificadores diferentes, RNA e MVS. Para ambos os classificadores foram testadas varias combinações ou topologias. Assim como em todas as experiencias realizadas foram usadas

duas doenças diferentes, como exemplos patológicos, separadas por género, obtendo assim na realidade quatro grupos patológicos.

Na tabela 12 podemos ver os melhores resultados alcançados para a RNA. Em média são necessários pelo menos dez neurónios na camada escondida para obter bons resultados para nove parâmetros na entrada. Existe uma supremacia da função de transferência da camada de saída *purelin* e na camada escondida *tansig*. Quanto á função de treino não existe uma função ótima para todos os casos.

Tabela 12-Melhores resultados da RNA para os parâmetros extraídos pelo Algoritmo apenas na vogal /a/ tom normal.

	Nº neur.	FTCE	FTS	FT	R-P2	Prec.-P2	R-T	Prec.-T
CFvsP40F	10	tansig	purelin	trainrp	0,257	62,2	0,707	83,3
CMvsP40M	10	tansig	purelin	trainlm	0,140	56,9	0,500	70,0
CFvsP136F	10	logsig	purelin	trainscg	0,404	69,0	0,358	65,8
CMvsP136M	5	tansig	purelin	trainlm	0,469	73,2	0,462	72,7

Na tabela 13 podemos observar os melhores resultados alcançados pela MVS. O *kernel* Gaussiano e o método *Quadratic Programming* parecem ser os que mais vezes permitem obter bons resultados.

Tabela 13-Melhores resultados da MVS para os parâmetros extraídos com o Algoritmo apenas na vogal /a/ tom normal. S=sigma, C=constante e O=ordem do polinómio.

	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
CFvsP40F	linear	C=0,1	QP	75,0	83,3	66, 7
CMvsP40M	Gauss.	S=0,1, C=0,1	QP	87,5	75,0	100
CFvsP136F	Gauss.	S=0,1, C=1	QP	76,3	84,2	68,4
CMvsP136M	Poli.	O=8, C=0,01	SMO	75,0	90,0	60,0

Na figura 10 é apresentado um comparativo entre os dois classificadores com os resultados obtidos utilizando os parâmetros extraídos pelo Algoritmo apenas para a vogal /a/ no tom normal. A sua análise permite-nos dizer que os melhores resultados são alcançados quando é utilizada a MVS. A MVS garante percentagens de precisão mais elevadas em três dos quatro grupos testados.

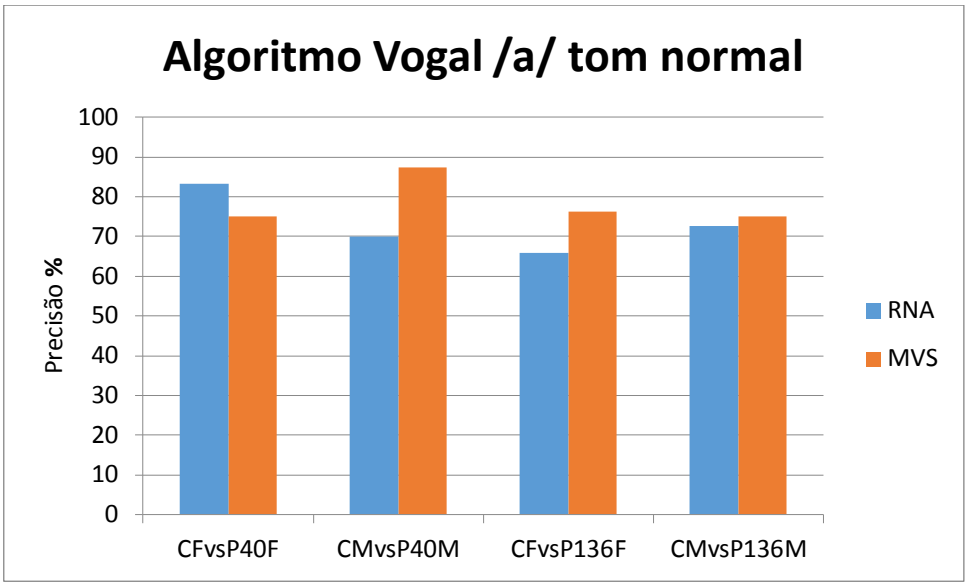

Figura 10-Comparativo entre classificadores para os parâmetros extraídos com o Algoritmo para a vogal /a/ tom normal.

6.1.2. CONCLUSÕES

A figura 11 representa os resultados dos classificadores (RNA e MVS) depois de aplicados os métodos de seleção de variáveis, no caso da utilização de varias vogais e tons, e os resultados para apenas a vogal /a/ no tom normal. A média é calculada para os quatro grupos utilizados: disfonia feminino, disfonia masculino, paralisia das cordas vocais feminino e paralisia das cordas vocais masculino.

Tanto na RNA como na MVS a aplicação dos métodos permitiram melhorar os resultados, tirando o caso do grupo da paralisia das cordas vocais feminino. O método 2 e PCA foram na maior parte dos casos melhores do que o método 1, por esse motivo e por o método 1 ser quase artesanal este foi excluído dos testes seguintes. Contudo, os melhores resultados alcançados nos grupos paralisia das cordas vocais feminino e masculino ficaram um pouco aquém do esperado, pelo que foram feitos esforços na procura de novos parâmetros que podem ser vistos na secção Experiencias com outros Parâmetros.

Globalmente a MVS e o método 2 permitem alcançar os melhores resultados. Contudo, volta a verificar-se que os resultados da paralisia das cordas vocais ficam um pouco abaixo do esperado. Tendo sido obtidos no caso feminino 78,9% e masculino 75%.

É possível ainda afirmar que os resultados melhoram com o uso de outras vogais e tons em relação ao uso de apenas uma vogal e um tom, como é o caso da vogal /a/ tom normal. Contudo, a patologia paralisia das cordas vocais continua um pouco aquém pelo que será necessário estudar outro tipo de parâmetros que melhorem estes resultados.

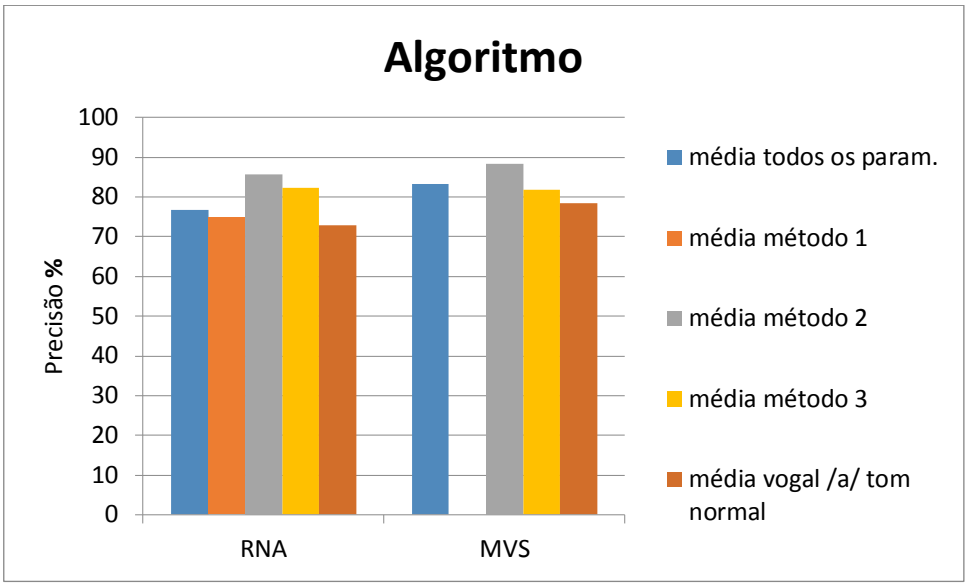

Figura 11-Comparativo entre métodos e classificadores para os parâmetros extraídos pelo algoritmo.

6.2. DESCRIÇÃO DAS EXPERIENCIAS COM OUTRO CONJUNTO DE PARÂMETROS

Nesta secção são relatadas as experiencias feitas com outros parâmetros já descritos na seção 4 e descrita a sua determinação na seção 5.

Foi desenvolvido um programa para extrair 12 coeficientes de cepstrais na frequência mel (MFCC), frequências e larguras de banda dos primeiros três formates (F1, F2, F3, Bw1, Bw2 e Bw3), frequência fundamental (F0), Energia, momentos espectrais de ordem 0, 1, 2, 3 e curtose (M0, M1, M2, M3, K) e a potência *root mean square* (RMS). Nesta experiência foi apenas utilizada a vogal /a/ no tom normal.

Foram treinados dois classificadores diferentes, RNA e MVS, utilizando como parâmetros de entrada todos os parâmetros e os modelos achados com o método 2 e 3 (PCA).

Foi testada a capacidade de predição dos classificadores para duas doenças diferentes, disfonia (P40) e paralisia das cordas vocais (P136). Os classificadores foram treinados com o mesmo número de exemplos positivos (Controlo) e negativos (Patológico). Foi feita a separação por género, ficando assim com controlo feminino e masculino (CF e CM), disfonia feminino e masculino (P40F e P40M) e paralisia feminino e masculino (P136F e P136M).

Foram testadas várias topologias e combinações tanto na RNA como MVS. Devido á inicialização aleatória dos pesos, na RNA foram usados 20 ciclos de treino e guardada a rede com melhor precisão com base no conjunto de validação. A seguir são apresentados os melhores resultados.

Foi ainda testada uma técnica de análise por *frames*. Esta técnica implica que em vez da média das *frames* sejam usados todos os *frames* disponíveis para treinar os classificadores. A divisão dos dados foi feita tendo em atenção quantos *frames* pertenciam a cada sinal para não haver parte das *frames* de um sinal no conjunto de validação e os restantes no conjunto de teste (por exemplo). O cálculo da precisão do conjunto de teste tinha por base que se determinada percentagem de *frames* (*threshold*) fossem atribuídos a uma dada classe então todos os *frames* desse sinal eram declarados como pertencentes a essa classe. Neste caso em vez dos 500 ms, foram utilizados os sinais depois de cortados os 100 ms do início e 100 ms do final do sinal, para obter o máximo de *frames* possível.

6.2.1. RESULTADOS

Nas tabelas 14, 15, 16 e 17 temos os melhores resultados para a RNA, com todos os parâmetros e com os modelos achados com a aplicação dos métodos 2 e 3 (PCA).

Para a classificação entre saudável e patológico foram usados 4 grupos diferentes, disfonia feminino (P40F, tabela 14), difonia masculino (P40M, tabela 15), paralisia das cordas vocais feminino (P136F, tabela 16) e paralisia das cordas vocais masculino (P136M, tabela 17).

Nas tabelas 14, 15, 16 e 17 temos informação sobre os parâmetros usados na entrada da rede (Entrada), o número de neurónios da camada escondida (Nº neur.) função de transferência da camada escondida (FTCE), função de transferência da saída (FTS), função de treino (FT), valor de R e Precisão do conjunto de treino mais validação (R-P2 e Prec.-P2) e valor de R e Precisão do conjunto de teste (R-T e Prec.-T).

Pela análise da tabela 14 vemos que a aplicação do método 2 proporcionou resultados iguais ao obtido com a utilização de todos os parâmetros, 91,7 %. O método 3 (PCA) reduziu a

precisão para 83,3%. Apesar de não haver melhorias estes resultados demonstram que apesar da redução do número de parâmetros não houve grandes perdas de informação.

Tabela 14-Resultados da RNA para a disfonia feminino (CFvsP40F).

Entrada	Nº neur.	FTCE	FTS	FT	R-P2	Prec.-P2 [%]	R-T	Prec.-T [%]
Todos param.	5	tansig	purelin	trainrp	0,439	72,0	0,845	91,7
Método 2-Modelo 13	5	logsig	purelin	trainrp	0,491	74,4	0,845	91,7
Método 3-PCA	5	logsig	purelin	trainscg	0,314	64,63	0,707	83,3

Na tabela 15 podemos observar que nem o método 2 nem o método 3 permitiram que os resultados obtidos com uso de todos os parâmetros fossem melhorados. Os métodos 2 e 3 obtiveram 80% e todos os parâmetros 90%.

Tabela 15-Resultados da RNA para a disfonia masculino (CMvsP40M).

Entrada	Nº neur.	FTCE	FTS	FT	R-P2	Prec.-P2 [%]	R-T	Prec.-T [%]
Todos param.	15	tansig	purelin	trainrp	0,140	56,9	0,816	90,0
Método 2-Modelo 14	5	tansig	purelin	trainscg	0,726	86,2	0,600	80,0
Método 3-PCA	20	tansig	purelin	trainlm	0,145	56,9	0,655	80,0

A tabela 16 demonstra uma clara melhoria dos resultados pela aplicação do método 2, 86,8%. Contudo o método 3 apenas permite um resultado idêntico ao uso de todos os parâmetros, 78,9%.

Tabela 16-Resultados da RNA para a paralisia das cordas vocais feminino (CFvsP136F).

Entrada	Nº neur.	FTCE	FTS	FT	R-P2	Prec.-P2 [%]	R-T	Prec.-T [%]
Todos param.	10	logsig	purelin	trainlm	0,604	80,2	0,610	78,9
Método 2-Modelo 15	5	logsig	purelin	trainrp	0,488	74,2	0,746	86,8
Método 3-PCA	5	tansig	purelin	trainlm	0,421	71,0	0,610	78,9

Na tabela 17 verifica-se um desempenho igual entre a utilização de todos os parâmetros e os modelos encontrados com o recurso aos métodos 2 e 3. Os valores de precisão alcançados para o conjunto de teste na patologia paralisia das cordas vocais masculino foi em ambos os casos de 85%.

Tabela 17-Resultados da RNA para a paralisia das cordas vocais masculino (CMvsP136M).

Entrada	Nº neur.	FTCE	FTS	FT	R-P2	Prec.-P2 [%]	R-T	Prec.-T [%]
Todos param.	10	logsig	purelin	trainlm	0,546	77,2	0,734	85,0
Método 2-Modelo 16	5	tansig	purelin	trainscg	0,633	81,6	0,734	85,0
Método 3-PCA	15	logsig	purelin	trainrp	0,454	72,1	0,704	85,0

Nas tabelas 18, 19, 20 e 21 podemos ver os melhores resultados alcançados para a MVS. Estas tabelas apresentam informação sobre o tipo de *kernel* utilizado (*kernel*), os parâmetros associados a esse *kernel* (Parâm.), o método de treino (Mét.), a precisão (Prec.-T), sensibilidade (Sens.) e especificidade (Espec.) do conjunto de teste.

A tabela 18 apresenta os melhores resultados da MVS para a disfonia feminino. Pela sua análise é possível verificar que os valores de precisão do conjunto de teste com o uso da MVS são iguais aos da RNA da tabela 14.

Tabela 18-Resultados da MVS para a disfonia feminino (CFvsP40F).

Entrada	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
Todos param.	linear	C=0,1	QP	91,7	83,3	100
Método 2-Modelo 13	linear	C=0,1	QP	91,7	83,3	100
Método 3-PCA	Gaussiano	S=1, C=0,1	SMO	83,3	83,3	83,3

A tabela 19 contém os resultados da MVS para o grupo disfonia masculino. A observação desta revela a não modificação dos valores de precisão quer pela aplicação do método 2 quer pelo método 3, 75%. Estes valores são contudo bastante inferiores aos registados com a RNA (tabela 15), 90%.

Tabela 19-Resultados da MVS para a disfonia masculino (CMvsP40M).

Entrada	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
Todos param.	Polinomial	O=6, C=0,1	QP	75,0	100	50,0
Método 2-Modelo 14	Polinomial	O=3, C=0,1	SMO	75,0	75,0	75,0
Método 3-PCA	linear	C=0.2	SMO	75,0	75,0	75,0

No caso da tabela 20, em que temos os resultados da MVS para a paralisia das cordas vocais feminino, podemos ver que os resultados são melhores do que os da RNA (tabela 16). Contudo existem semelhanças, a precisão aumenta com o uso do método 2 e mantem-se com o uso de todos os parâmetros e método 3.

Tabela 20-Resultados da MVS para a paralisia das cordas vocais feminino (CFvsP136F).

Entrada	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
Todos param.	Gaussiano	S=4, C=0,2	SMO	81,6	89,5	73,7
Método 2-Modelo 15	Polinomial	O=2, C=0,01	SMO	84,2	89,5	78,9
Método 3-PCA	Gaussiano	S=3, C=1	SMO	81,6	89,5	73,7

Na tabela 21 podemos observar os resultados obtidos com a MVS para a paralisia das cordas vocais masculino. A sua analise permite-nos constatar que os valores de precisão são em tudo similares aos obtidos com a RNA (tabela 17) com a exceção da aplicação do método 2 que permitiu uma melhoria de 85% para 90%.

Tabela 21-Resultados da MVS para a paralisia das cordas vocais masculino (CMvsP136M).

Entrada	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
Todos param.	linear	C=0,1	QP	85,0	70,0	100
Método 2-Modelo 16	linear	C=0,01	QP	90,0	80,0	100
Método 3-PCA	linear	C=0,01	SMO	85,0	100	70,0

Na tabela 22 podemos ver os resultados alcançados pela RNA com recurso á técnica de análise por *frames*. Nela constam dados sobre as topologias que permitiram obter os melhores resultados como o número de neurónios (Nº neur.), função de transferência da camada escondida (FTCE), função de transferência da saída (FTS) e função de treino (FT).

São apresentados os valores de precisão do conjunto de teste (Prec.-T) para as duas doenças separadas por género com recurso a um *threshold* de 70% e de 50%.

Como podemos observar os melhores resultados são obtidos usando um *threshold* de 50%. Este valor pode levantar algumas questões por ser o valor mínimo admissível para se considerar que um sinal pertence a determinada classe. O uso de um *threshold* de 70% permite ainda assim obter bons resultados nos grupos do género feminino.

Tabela 22-Resultados da RNA para o uso da técnica de analise por frames, com threshold a 50% e 70%.

		Nº neur.	FTCE	FTS	FT	Prec.-T [%]
Threshold 50%	CFvsP40F	5	tansig	purelin	trainlm	100
	CMvsP40M	25	logsig	purelin	trainrp	100
	CFvsP136F	20	logsig	purelin	trainrp	100
	CMvsP136M	20	tansig	purelin	trainscg	85
Threshold 70%	CFvsP40F	5	tansig	purelin	trainlm	100
	CMvsP40M	5	logsig	purelin	trainlm	70
	CFvsP136F	15	tansig	purelin	trainrp	97,4
	CMvsP136M	10	tansig	purelin	trainscg	80

Aqui só são apresentados os resultados da RNA porque não foi possível testar todas as topologias com a MVS. A MVS tem como opção vários ajustes para procurar melhorar os resultados. Contudo, devido ao acréscimo do número de exemplos de treino, fruto da análise por *frames*, não foi possível usar alguns destes recursos. Nomeadamente o método *quadratic programming* (QP) tornou-se de tal maneira exigente do ponto de vista computacional que esgotou toda a RAM do computador usado nos testes. Como tal, para que houvesse uma igualdade de critérios apenas foi usada a RNA.

6.2.2. CONCLUSÕES

A figura 12 representa os resultados dos classificadores (RNA e MVS) depois de aplicados os métodos de seleção de variáveis. A média é calculada para os quatro grupos utilizados: disfonia feminino, disfonia masculino, paralisia das cordas vocais feminino e paralisia das cordas vocais masculino. A aplicação do método 2, regra geral, ou mantém os resultados obtidos ou melhora. Contudo, a exceção á regra é o caso em que a aplicação deste método no grupo disfonia masculino reduziu a precisão de 90% para 80%. Globalmente o método 2 permite melhorar os resultados ao mesmo tempo que reduz a dimensão do problema. Contudo, os melhores resultados são obtidos com o uso de todos os parâmetros e a RNA (figura 12). Se recorrermos á técnica de análise por *frames* os resultados melhoram significativamente. Principalmente se for usado um *threshold* de 50%. A utilização de um *threshold* de 70% consegue ainda assim alcançar bons resultados nos grupos do género feminino. Contudo, se pretendermos uma precisão elevada para todos os grupos é necessário baixar este *threshold* para 50%. Embora possa suscitar algumas dúvidas por este valor ser o mínimo admissível para se considerar um sinal como pertencente a determinada classe, o seu uso não é de todo descabido.

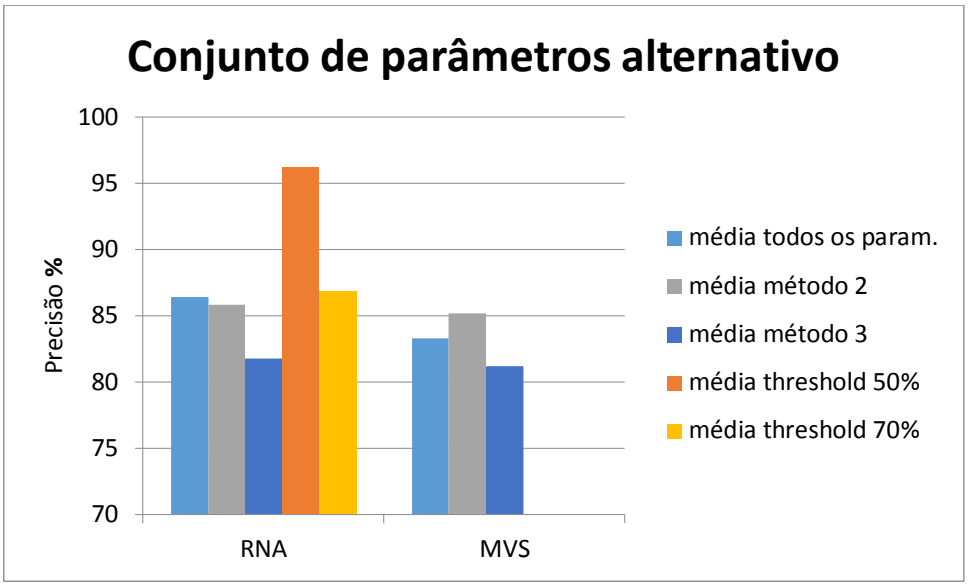

Figura 12-Comparativo entre métodos e classificadores para o conjunto de parâmetros alternativo.

6.3. DESCRIÇÃO DAS EXPERIENCIAS COM O PRAAT

O Praat foi usado para determinar os conjuntos de parâmetros 1 e 2. O objetivo seria o de verificar se haveria diferenças significativas no diagnóstico usando os parâmetros determinados de forma diferente, já que os valores dos parâmetros obtidos pelo Praat e pelos métodos descritos atrás nem sempre são iguais.

Na primeira experiencia os scripts escritos para correr no Praat permitiam extrair 12 coeficientes cepstrais na frequência mel (MFCC), frequências e larguras de banda dos primeiros três formates (F1, F2, F3, Bw1, Bw2 e Bw3), frequência fundamental ou *Frequência fundamental* (F0), Energia, momentos espectrais de ordem 0, 1, 2, 3 e curtose (M0, M1, M2, M3, K) e a potência. Nesta experiência foi apenas utilizada a vogal /a/ no tom normal. Por se tratar um conjunto de parâmetros ainda grande foram aplicadas as técnicas de seleção de variáveis já descritas anteriormente. Estes parâmetros correspondem ao conjunto de parâmetros 2.

Na segunda Experiencia os scripts escritos no Praat permitiam extrair jitter absoluto, jitter relativo, jitter ppq5, jitter rap, shimmer absoluto, shimmer relativo, shimmer apq3, shimmer apq5 e HNR. Também estes apenas na vogal /a/ e tom normal. Estes parâmetros correspondem ao conjunto de parâmetros 1.

6.3.1. RESULTADOS

Na tabela 23 podemos ver os melhores resultados alcançados por dois classificadores diferentes (RNA e MVS), para quatro grupos patológicos (disfonia feminino e masculino, paralisia das cordas vocais feminino e masculino) usando o conjunto de parâmetros 2 assim como a aplicação de técnicas/métodos de seleção de variáveis. A Precisão refere-se ao conjunto de teste apenas (Prec.-T). A tabela 23 é uma tabela resumo com os melhores resultados, tendo sido testadas varias combinações e topologias.

Pela análise da tabela 23 podemos afirmar que a utilização do método 2, na maior parte dos casos, permite melhorar ou manter os valores de precisão. O método 3 (PCA) consegue também melhorar em grande parte dos casos os resultados contudo não supera o método 2.

O uso da MVS aliado ao método 2 é o sistema que garante os melhores resultados.

Tabela 23-Resultados usando o conjunto de parâmetros 2 determinados com o Praat para a RNA e MVS.

		Todos os param. Prec.-T [%]	Método 2 Prec.-T [%]	Método 3 Prec.-T [%]
RNA	CFvsP40F	83,3	100	100
	CMvsP40M	90	90	80
	CFvsP136F	76,3	76,3	78,9
	CMvsP136M	85	100	95
MVS	CFvsP40F	83,3	100	91,7
	CMvsP40M	87,5	100	87,5
	CFvsP136F	76,3	81,6	71,1
	CMvsP136M	100	95	95

Nas tabelas 24 e 25 podemos observar os resultados obtidos para o conjunto de parâmetros 2.

A tabela 24 refere-se aos melhores resultados alcançados pela RNA para o conjunto de parâmetros 2 extraídos com o Praat. Como podemos ver pela análise da tabela, os melhores resultados são obtidos em grande parte usando apenas cinco neurónios na camada escondida para nove parâmetros de entrada. São nove parâmetros de entrada porque são usados o conjunto de parâmetros 1 apenas na vogal /a/ tom normal. A função de transferência *logsig* aparece associada aos casos femininos enquanto que a *tansig* aparece associada aos masculinos. Parece haver também uma tendência para obter melhores resultados no género masculino de cada patologia.

Tabela 24-Melhores resultados da RNA para o conjunto parâmetros 1 extraídos com o Praat.

	Nº neur.	FTCE	FTS	FT	R-P2	Prec.-P2	R-T	Prec.-T
CFvsP40F	10	logsig	purelin	trainlm	0,688	84,1	0,507	75,0
CMvsP40M	5	tansig	purelin	trainscg	0,466	72,4	0,816	90,0
CFvsP136F	5	logsig	purelin	trainscg	0,388	69,4	0,639	81,6
CMvsP136M	5	tansig	purelin	trainrp	0,588	78,7	0,800	90,0

A tabela 25 contém os melhores resultados obtidos com a MVS para o segundo conjunto de parâmetros extraídos com o Praat. A análise desta permite verificar uma tendência para o uso de um *kernel* linear com método *Quadratic Programming*. O uso da MVS em vez da RNA fez com que houvesse uma ligeira melhoria nos resultados de uma forma geral.

Tabela 25-Melhores resultados da MVS para conjunto de parâmetros 1 extraídos com o Praat. S=sigma, C=constante e O=ordem do polinómio.

	Kernel	Parâm.	Mét.	Prec.-T	Sens.	Espec.
CFvsP40F	linear	C=10	QP	87,5	100	75,0
CMvsP40M	linear	C=0,1	QP	87,5	100	75,0
CFvsP136F	Gauss.	S=0,1, C=0,1	QP	81,6	79,0	84,2
CMvsP136M	linear	C=0,01	LS	90,0	90,0	90,0

6.3.2. CONCLUSÕES

Na Figura 13 constam as médias da precisão ao longo dos quatro grupos patológicos para o conjunto de parâmetros 1 e 2 extraídos pelo Praat usando RNA e MVS. O uso do conjunto de parâmetros 1 e 2 obtém resultados idênticos não havendo vantagem significativa em usar um conjunto ou outro. Por seu turno a aplicação do método 2 ao conjunto de parâmetros 2 fez com que houvesse uma melhoria significativa nos resultados. Pode-se concluir que a aplicação do método 2 (regressão linear passo a passo) permite melhorar os resultados reduzindo a dimensão dos parâmetros de entrada. Por outro lado a hipótese de se juntar os dois conjuntos de parâmetros não foi aqui testada e contínua em aberto, podendo vir esta simbiose a beneficiar da aplicação do método 2 que demonstrou grande potencial.

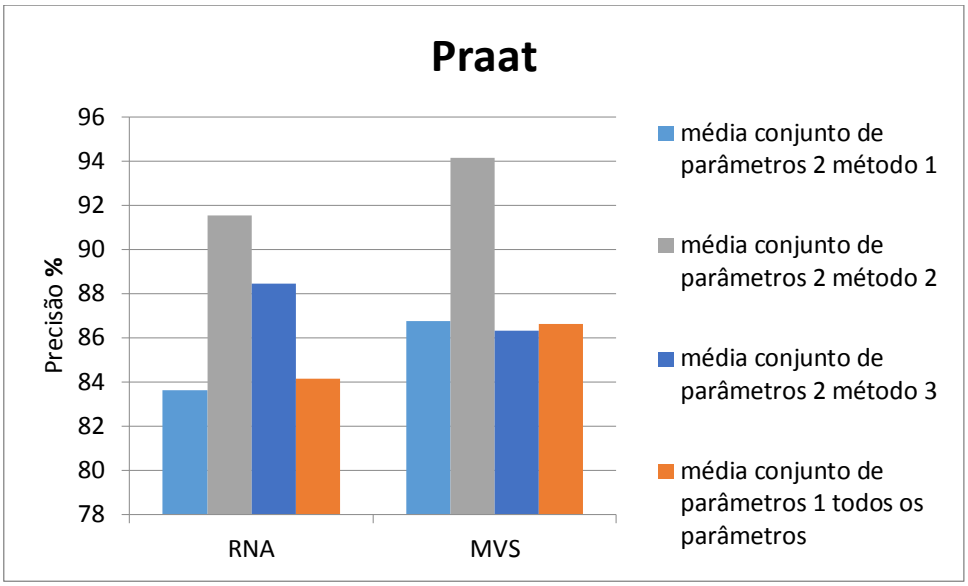

Figura 13- Comparativo entre métodos e classificadores para o conjunto de parâmetros 1 e 2 extraídos com o Praat.

6.4. DISCUSSÃO

Nesta secção vão ser comparados os vários conjuntos de parâmetros, métodos de seleção de parâmetros e classificadores usados por forma a aferir qual o melhor sistema de diagnóstico.

Na figura 14 podem ser observados os melhores resultados obtidos por algoritmo, independentemente do método e classificador, para cada grupo patológico. Os métodos de seleção de variáveis e classificadores usados podem ser vistos na tabela 26.

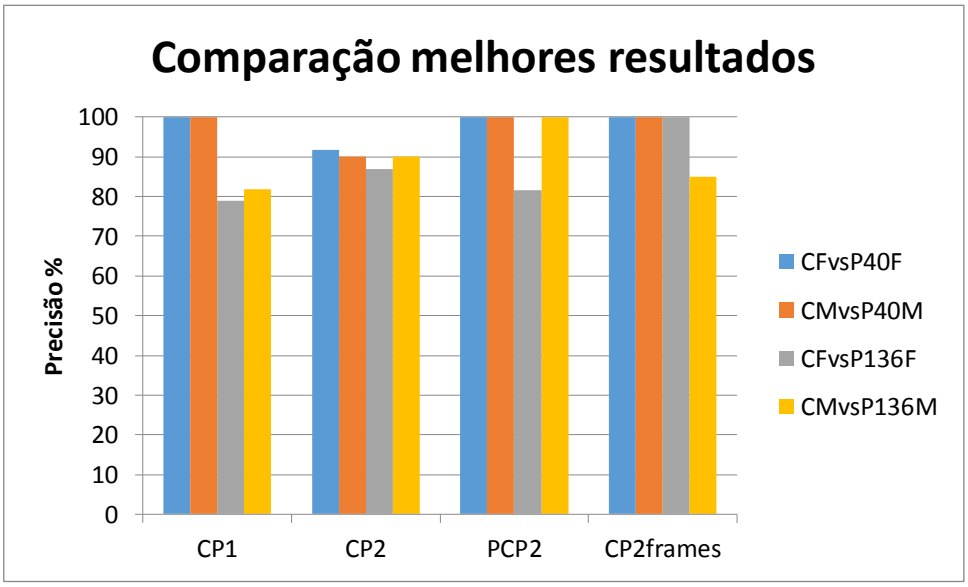

Figura 14-Comparativo entre os melhores resultados obtidos por algoritmo para cada grupo patológico, independentemente do método ou classificador.

Os resultados assinalados como CP1 correspondem ao conjunto de parâmetros 1 extraídos com o algoritmo desenvolvido por Gonçalves. Este conjunto de parâmetros é composto por quatro parâmetros de Jitter, quatro parâmetros de Shimmer e HNR para três vogais e três tons diferentes. Os modelos obtidos por aplicação dos métodos de seleção de variáveis aliados a determinado classificador permitiram melhorar a performance e alcançar os valores de precisão do conjunto de teste que são observados na figura 14. Os resultados obtidos para a disfonia, tanto no caso feminino como masculino são de 100%. Contudo a patologia paralisia das cordas vocais tanto num género como no outro não alcançaram valores desta ordem de grandeza. Ficando-se pelos 78,9% e 81,8%. Este facto levou a que fossem procurados outros parâmetros no sentido de melhorar estes resultados.

O conjunto de parâmetros 2 (CP2) foram extraídos apenas para a vogal /a/ no tom normal. No caso do CP2 estes foram submetidos á aplicação das mesmas técnicas de seleção de variáveis que CP1, enquanto que em CP2frames foi aplicada uma técnica de análise por *frames* usando um *threshold* de 50%. A aplicação desta técnica só é possível porque este conjunto de parâmetros são na sua maioria parâmetros de análise de curto termo, que pressupõe que as características do sinal permaneçam invariáveis por um período curto tempo da ordem dos 20 a 40 ms. Como podemos observar na figura 14 o uso deste conjunto de parâmetros permitiu melhorar os resultados da patologia paralisia das cordas vocais, em relação a CP1, tanto no género feminino como masculino, 86,8% e 90% respetivamente. Contudo, no caso da disfonia estes resultados foram inferiores ao anterior. Ainda assim obtiveram 91,7% no feminino e 90% no masculino. Quanto á técnica de análise por *frames* (CP2frames), se for usado um *threshold* de 50%, os resultados melhoram em quase todos os grupos para 100%, excetuando a paralisia das cordas vocais masculino que passa de 90% para 85%.

Os resultados foram também comparados usando os parâmetros determinados pelo Praat. Este programa já foi usado anteriormente, nomeadamente na validação do algoritmo desenvolvido em (Teixeira & Gonçalves, 2014, e Teixeira & Gonçalves, 2016). Foram feitos vários testes com pelo menos dois conjuntos de parâmetros distintos, um conjunto de parâmetros iguais aos CP1 mas apenas para a vogal /a/ e tom normal e um segundo conjunto de parâmetros semelhantes aos CP2, também apenas para a vogal /a/ e tom normal. Na figura 14 apenas estão os resultados do segundo conjunto de parâmetros extraídos com o Praat. Sendo assim os resultados assinalados como PCP2 correspondem a um conjunto de parâmetros extraídos pelo Praat. Estes parâmetros são em tudo semelhantes aos CP2. Como podemos ver na figura 14 este conjunto de parâmetros consegue alcançar valores de precisão melhores do que os CP2

tirando o caso da paralisia das cordas vocais feminino em que CP2 tem 86,8% e PCP2 tem 81,6%. Se forem comparados com CP2frames, os PCP2 levam a melhor na paralisia masculino mas CP2frames tem melhores resultados na paralisia feminino. Contudo, a média de resultados é superior em CP2frames.

O método que mais vezes garante bons resultados é o método 2 (regressão linear passo a passo), como podemos observar na tabela 26. Quanto ao classificador existe uma ligeira vantagem na utilização das RNA. Devido á inicialização aleatória dos pesos a RNA necessita de mais ciclos de treino para obter um resultado favorável. No caso das MVS o resultado é sempre o mesmo após treino. Apesar dessa ligeira desvantagem que pode tornar a obtenção de resultados mais tardia vale a pena usar as RNA.

Tabela 26-Tabela auxiliar á figura 14 com os métodos e classificador para cada algoritmo e grupo patológico.

	CP1	CP2	PCP2
CFvsP40F	Método 2-RNA	Método 2-RNA	Método 2-RNA
CMvsP40M	Método 2-MVS	Todos Parâmetros-RNA	Método 2-MVS
CFvsP136F	Método 2-MVS	Método 2-RNA	Método 2-MVS
CFvsP136M	Método 2-RNA	Método 2-MVS	Método 2-RNA

CAPÍTULO 7

7. CONCLUSÕES E TRABALHOS FUTUROS

7.1. CONCLUSÕES

Tendo em conta as experiências realizadas, pode-se concluir que o uso dos quatro parâmetros de jitter, shimmer e HNR, apenas permitem obter bons resultados para uma das doenças, disfonia, devido ao uso de várias vogais e vários tons. Por si só este conjunto de parâmetros não apresenta grande poder preditivo para apenas uma vogal e tom. Utilizando um segundo conjunto de parâmetros, ainda que apenas na vogal /a/ e tom normal, é possível obter melhores resultados na paralisia das cordas vocais do que o obtido com o jitter, shimmer e HNR para várias vogais e vários tons. Contudo, o jitter, shimmer e HNR para várias vogais e tons continua a ser a melhor forma de classificar como patológico ou saudável quando se usa a disfonia. Os resultados obtidos usando o segundo conjunto de parâmetros indiciam que é possível ainda melhorar e explorar outro tipo de parâmetros.

Existe contudo a possibilidade já testada, e que parece apresentar bons resultados, de utilizar uma técnica de análise por *frames*. Esta técnica é possível de ser aplicada devido ao conjunto de parâmetros 2 serem parâmetros de análise de curto termo. Isto pressupõe que o sinal tem características de estacionaridade num curto período de tempo da ordem dos 20 a 40 ms. Como tal, esta técnica permite avaliar a variação de determinada grandeza física num período curto de tempo trazendo assim vantagens em relação a técnicas em que usam um período mais longo para calcular uma grandeza física. A outra opção seria aliar os parâmetros de jitter, shimmer e HNR a este conjunto de parâmetros de curto termo. Mas apenas usando o jitter absoluto e shimmer relativo uma vez que os restantes como se veio a verificar, estão bastante correlacionados entre si.

Como podemos verificar tanto nas experiencias relatadas com o Praat como nesta ultima comparação, os parâmetros extraídos com o Praat parecem apresentar um poder classificativo superior aos parâmetros determinados de outras formas descritas.

Quanto á técnica de seleção de variáveis existe uma quase hegemonia por parte da regressão linear passo a passo com exceção de CP2 em que no grupo disfonia masculino os melhores

resultados são obtidos utilizando todos os parâmetros. O classificador com melhores resultados em média é a Rede Neuronal Artificial (RNA).

No trabalho desenvolvido em Fezari et al, 2014 os melhores resultados obtidos usando a mesma base de dados (SVD) e a patologia disfonia espasmódica foram de 82,3%. Já em Panek et al, 2015 usando a base de dados SVD e a patologia paralisia das cordas vocais é obtido uma precisão de 100%. No trabalho desenvolvido, na classificação como saudável ou patológico, foram obtidos valores de precisão de 100% quando usado o conjunto de parâmetros 1 (CP1) na disfonia. O conjunto de parâmetros 2 (CP2) permitiu obter uma precisão de 90% na paralisia das cordas vocais masculino e usando a técnica de *frames* (CP2frames) de 100% na paralisia das cordas vocais feminino.

Devido á reduzida dimensão do conjunto de dados de teste estes resultados devem ser relativizados uma vez que podem não ter valor estatístico significativo.

7.2. TRABALHOS FUTUROS

Como trabalhos futuros gostaria de sugerir que fossem usadas outras bases de dados para aumentar a dimensão dos conjuntos de teste de forma a obter resultados com maior significado estatístico.

Como se verificou existe um grande potencial por parte dos dois conjuntos de parâmetros quando usados de forma individual. Como tal a combinação dos conjuntos de parâmetros 1 e 2 pode eventualmente melhorar os resultados obtidos.

A classificação entre saudável e patológico pode também ser feita usando várias outras patologias. Pode ser criado um grupo patológico que contenha muitas mais patologias do aqui foram usadas. Podendo numa primeira fase ser tentada a classificação como patológico e saudável e numa segunda fase identificar o tipo de patologia.

Será necessário também fazer uma validação do conjunto de parâmetros 2 uma vez que estes apresentam valores relativamente diferentes dos determinados pelo Praat.

BIBLIOGRAFIA

Almeida, N. C. Sistema Inteligente para Diagnóstico de Patologias na Laringe Utilizando Máquinas de Vetor de Suporte. Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, 2010.

Al-nasheri, A., Muhammad, G., Alsulaiman, M., Ali, Z. Investigation of Voice Pathology Detection and Classification on Different Frequency Regions Using Correlation Functions. Journal of Voice, 2016.

Ben-Hur, A., Weston, J. A User's Guide to Support Vector Machines. Methods in molecular biology (Clifton, N.J.), 609, pp. 223–239, 2010.

Bishop, C. M. Neural Network for Pattern Recognition.Clarendon Press , Oxford, 1995.

Boersma, P., Heuven, V. Speack and Unspeack With Praat. Glot International Vol. 5, No. 9/10, (341-347), Blackwell Publishers Ltd. 2001.

Brockmann-Bauser, M. Improving jitter and shimmer measurements in normal voices. Institute of Cellular Medicine, Medical School, Newcastle University, 2011.

Catford, J. C. A Pratical Introduction to Phonetics, Sec. Ed., Oxford University Press, 2001.

Cordeiro, H. T., Fonseca, J. M., Ribeiro, C. M. LPC Spectrum First Peak Analysis for Voice Pathology Detection. Procedia Technology, 2013.

Cruz, A. J. R. Data Mining via Redes Neuronais Artificiais e Máquinas de Vectores de Suporte. Escola de Engenharia, Universidade do Minho, 2007.

Draper, N. R., Smith, H. Applied Regression Analysis, Third Edition. Wiley Series in Probability and Statistics, 1998.

Eskidere, O., Gurhanh, A. Voice Disorder Classification Based on Multitaper Mel Frequency Cepstral Coefficients Features. Hindawi Publishing Corporation, Computational and Mathematical Methods in Medicine, 2015.

Fezari, M., Amara, F., M. M. El-Emary, I. Acoustic Analysis for Detection of Voice Disorders Using Adaptive Features and Classifiers. International Conference on Circuits, Systems and Control, 2014.

Forero, L. A., Kohler, M., Vellasco, M., Cataldo, E. Analysis and Classification of Voice Pathologies Using Glottal Signal Parameters. Journal of Voice, 2015.

Fujinaga, I. Adaptative Optical Music. McGuill University, Montreal, Canada, 1996.

Godino-Llorente, J. I., Aguilera-Navarro, S., Gómez-Vilda, P. LPC, LPCC And MFCC Parameterisation Applied to The Detection of Voice Impairments. Sixth International Conference on Spoken Language Processing, ICSLP 2000 / INTERSPEECH, 2000.

Guyon, I., Elisseeff, A. An Introduction to Variable and Feature Selection- Journal of Machine Learning Research 3, 1157-1182, 2003.

Henríquez, P., Alonso, J. B., Ferrer, M. A., Travieso, C. M., Godino-Llorente, J. I., Díaz-di-María, F. Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 17, No. 6, August 2009.

Kotsiantis, S. B. Supervised machine learning: a review of classification techniques. Informatica, 31, pp. 249–268, 2007.

Lanc, T. L. The Importance of Input Variables to a Neural Network Fault-diagnostic System for Nuclear Power Plants. Retrospective Theses and Dissertations, Paper 208, Iowa, 1992.

Logan, B. Mel Frequency Cepstral Coeffiecients for Music Modeling. International Symposium on Music Information Retrieval, Cambrige Research Laboratory, 2000.

Lopes, J. M. Ambiente de Análise Robusta dos Principais Parâmetros Qualitativos da Voz. Faculdade de Engenharia da Universidade do Porto, 2008.

Malyska, N., Quatieri, T. F., Sturim, D. Automatic Dysphonia Recognition Using Biologically-Inspired Amplitude-Modulation Features. IEEE, 2005.

Markaki, M., Stylianou, Y. Voice Pathology Detection and Discrimination Based on Modulation Spectral Features. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 19, No. 7, 2011.

Mathworks, Community, File Exchange, https://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab. Consultado pela ultima vez em: 26/09/2016.

Mathworks, Support, http://www.mathworks.com/help/nnet/ug/divide-data-for-optimal-neural-network-training.html?searchHighlight=neural%20network%20data%20division. Consultado pela última vez em: 24/09/2016.

Mathworks, Support, http://www.mathworks.com/help/signal/ug/formant-estimation-with-lpc-coefficients.html , Consultado pela ultima vez em 21/09/2016.

May, R., Dandy, G., Maier, H. Review of Input Variable Selection Methods for Artificial Neural Networks-Methodological Advances and Biomedical Applications, Prof. Kenji Suzuki (Ed.), 2011.

Molau, S., Pitz, M., Schulter, R., Ney, H. Computing Mel Frequency Cepstral Coefficients on the Power Spectrum. Acoustics, Speech, and Signal Processing, IEEE International Conference, 2001.

Moraes, R., Valiati, J. F., Neto, W. P. G. Documente-level Sentiment classification: An Impirical Comparison Between MVS and ANN. Expert Systems With Applications, 621-633, Elsevier, 2013.

Moran, J. R., Reilly, R. B., Chazal, P., Lacy, P. Telephony-Based Voice Pathology Assessment Using Automated Speech Analysis. IEEE Transactions On Biomedical Engineering, Vol. 53, No. 3, 2006.

Muda, L., Begam, M., Elamvazuthi, I. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. JOURNAL OF COMPUTING, Vol. 2, 2010.

Panek, D., Skalski, A., Gajda, J., Tadeusiewicz, R. Acoustic Analysis Assessment in Speech Pathology Detection. Int. J. Appl. Math. Comput. Sci., 2015, Vol. 25, No. 3, 631–643.

Paul Boersma, Manual Praat, 2003.

Poomjan, S., Taengtang, T., Srinuanjan, K., Kamoldilok, S., Ruttanapun, C., Buranasiri, P. Proof of Using Fourier Coefficients for Root Mean Square Calculations on Periodic Signals. Adv. Studies Theor. Phys., Vol. 8, No. 1, 21 – 25, 2014.

Pylypowich, A., Duff, E. Differentiating the Symptom of Dysphonia. The Journal for Nurse Practitioners. Elsevier, 2016.

Rawlings, J. O., Pantula, S. G., Dickey, D. A. Applied Regression Analysis: A Research Tool, Second Edition. Springer, 1998.

Rodriguez, M. Simultaneous Regression and Clustering to Predict Movie Ratings. Tese de Mestrado, University of California, 2010.

Salhi, L., Mourad, T., Cherif, A. Voice Disorders Identification Using Multilayer Neural Network. The International Arab Journal of Information Technology, Vol. 7, No. 2, 2010.

Schwarz, D. Spectral Envelopes in Sound Analysis and Synthesis. IRCAM, Institudo para a Informática, Estugarda, 1998.

Sellam, V., Jagadeesan, J. Classification of Normal and Pathological Voice Using MVS and RBFNN. Journal of Signal and Information Processing, 2014, 5, 1-7

Smith, L. I. A tutorial on Principal Components Analysis ,2002.

Sweitzer, K. A., Bishop, N. W. M., Genberg, V. L. Efficient Computation of Spectral Moments for Determination of Random Response Statistics. Proceedings of ISMA, 2004.

Tan, L., Karnjanadecha, M. Frequência fundamental Detection Algorithm: Autocorrelation Method and AMDF. Proceedings of the 3rd International Symposium on Communications and Information Technology, vol. 2, pp. 551–556, 2003.

Teixeira, J. P., Fernandes, P. O. Acoustic Analysis of Vocal Dysphonia. Procedia Computer Science. Elsevier, 2015.

Teixeira, J. P., Gonçalves, A. "Algorithm for jitter and shimmer measurement in pathologic voices", Procedia Computer Science - Elsevier 100 (2016) 271 – 279.

Teixeira, J. P., Gonçalves, A. Accuracy of Jitter and Shimmer Measurements. Procedia Technology. Elsevier, 2014.

Tiwari, V. MFCC and its applications in speaker recognition. International Journal on Emerging Technologies, 2010

U.S. Department of Health & Human Services, National Institute on Deafness and Other Communication Disorders (NIDCD). NIDCD Fact Sheet: Vocal Fold Paralysis, Publication No. 11-4306, 2011. https://www.nidcd.nih.gov/health/vocal-fold-paralysis

Uloza, V., Verika, A., Bacauskiene, M., Gelzinis, A., Pribuisiene, R., Kaseta, M., Saferis, V. Categorizing Normal and Pathological Voices:Automated and Perceptual Categorization. Journal of Voice, Vol. 25, No. 6, pp. 700-708, 2010.

Vogel, F., Holm, S., Lingdjaerd, O. C. Spectral Moments and Time Domain Representation of Photoacustic Signals Used for Detection of Crude Oil in Produced Water. Universidade de Oslo, Noruega, 2001.

Zekic-Susac, M., Sarlija, N., Pfeifer, S. Combining PCA Analysis and Artificial Neural Networks in Modelling Entrepreneurial Intentions of Students. Croatian Operational Research Review (CRORR), Vol. 4, 2013.

Zhang, G. P. Neural Networks for Classification: A Survey. IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, Vol. 30, No. 4, 2000.